ROLE OF CONFORMATIONAL CHANGES IN G PROTEIN-COUPLED RECEPTOR ACTIVATION

Thesis by
Caitlin E. Scott

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California
2014
(Defended May 23, 2014)

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor William A. Goddard III, for his guidance for the past seven years. He is extremely creative and imaginative and a good source for new ideas. He has allowed me the freedom to pursue interesting projects. He has supported me and provided me with great advice as well as challenged me to do the best work possible. He always was confident in my abilities.

Second, I would like to thank Professor Ravinder Abrol for his tremendous guidance and patience. He was always willing to discuss science and be supportive. I wish him the best of luck as a faculty member. His future postdocs and graduate students will be lucky to work with him.

Third, I would like to thank Professor Debra A. Kendall and Dr. Kwang H. Ahn, our experimental collaborators at the University of Connecticut. Without them, this work would not have been possible. They have been supportive and patient colleagues. I thank them for all their help and advice.

I would also like to thank our other experimental collaborators including Dr. Prasadarao V. Nemani of Children's Hospital Los Angeles and the Keck School of Medicine of the University of Southern California and Dr. John T. Groves of Princeton University. Dr. Nemani, who worked with us on the Ecgp project, is very creative, intelligent, and enthusiastic about his students' research. Dr. Groves and his lab, who worked with us on my vanadium catalysis project, do outstanding and high quality research.

Next, I would like to acknowledge the various members of the Goddard lab. They are all very friendly and intelligent. Many of them are willing to take hours out of their days to answer questions, demonstrate software, or explain concepts. I was lucky to be a part of this group. I focused on computational applications, which would not have been possible without the programmers and method developers—Adam Griffith, Dr. Jenelle

Bray, Dr. Ismet Caglar Tanrikulu, Dr. Bartosz Trzaskowski, Dr. Tod Pascal, and Vaclav Cvicek. I would also like to thank the undergraduate students who helped me with my work—Yuehan Huang, Nicholas Parker, and Kyle Tejada. They are intelligent and talented scientists who should go far in their respective careers whether they continue science or not. I also owe a lot to the following group members who provided immense amounts of feedback and advice: Dr. Soo-Kyung Kim, Andrea Kirkpatrick, Fan Liu, Matthew Gethers, Dr. William Ford, Dr. Henry Levenson, Dr. Heather Wiencko, Samantha Johnson, Dr. Wei-Guang Liu, Dr. Ross Fu, Dr. Himanshu Mishra, and Yan Choi Lam. Special thanks to Dr. Robert 'Smith' Nielsen for teaching me about catalysis and being so patient.

I would like to thank the members of my thesis committee: Professor Harry Gray, Professor William Clemons, and Professor Shu-ou Shan. They provided valuable critiques and suggestions that made me a better scientist.

Lastly, I would like to thank my family and many friends for their support and encouragement. I couldn't have done it without you.

My graduate work at Caltech was initially funded by PharmSelex/Accelerator and more recently by the NIH grants R01NS071112, R01NS073115, R01AI040567 and Sanofi. My final year of graduate school was funded by the Center for Catalytic Hydrocarbon Functionalization (CCHF) and the Bing Scholarship from the Caltech Dean's Office. The Materials and Process Simulation Center (MSC) and Caltech teaching fellowships have provided continuous support.

ABSTRACT

Transmembrane signal transduction is achieved by activation of G protein-coupled receptors (GPCRs) like the human cannabinoid type 1 (CB1) receptor, the human cannabinoid type 2 (CB2) receptor, and the human mu-opioid receptor. These receptors exist in the membrane in an ensemble of conformations each of which might bind to different signaling molecules and cause different physiological effects. Understanding the structural basis of their activation will eventually help us in designing drugs that target these receptors with potentially minimal undesirable side effects. CB1 is of particular interest because it is located in the central nervous system and modulates hunger, making it an attractive anti-obesity drug target. In this receptor, mutating a single residue, threonine 210, to isoleucine in the third transmembrane (TM3) domain makes it far more active than the wild-type (WT) receptor, whereas mutating it to alanine makes it fully inactive. CB1 is difficult to model because it has a small sequence identity with the receptors that have been crystallized. We used the first principles-based GEnSeMBLE method to predict 3D structures of these receptors representing the fully inactive to highly constitutively active states. With this software, we quickly found a set of low energy receptor conformations by sampling trillions of helix orientations. Differences in the intracellular surface explain experimental differences in activation for the CB1 receptor and its mutants. These predictions were validated by designing double mutants that were expected to switch the inactive T210A to WT levels of activation and expected to switch the very active L207A to T210A levels of activation. These predictions were first verified computationally then experimentally with GTPγS assays. The accuracy of our predictions indicate that the GEnSeMBLE method is a useful procedure for predicting GPCR structures at various activation states. Known inverse agonists were docked to these predicted CB1 receptor structures, and the resulting complexes were inserted into a solvated lipid bilayer for 50 ns of NPT molecular dynamics with NAMD software. The inverse agonist preferentially binds to a pre-activated CB1 state, but during MD, traits of the inactive structure start to form suggesting that the ligand induces conformational changes.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter VI Figures

Chapter VII Figures

Chapter VIII Figures

Chapter IX Figures

LIST OF TABLES

Chapter IX Tables

Chapter I

INTRODUCTION

Adapted with permission from:
Scott CE, Abrol R, Ahn KH, Kendall DA, Goddard WA, III (2013) Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations. Protein Sci **22**:101-113.

BACKGROUND

Transmembrane proteins, existing on the cellular surface, are important drug targets due to their ability to relay information from the environment to the cell interior. We are interested in two types of drugs—the ones that interact with transmembrane proteins to cause downfield signaling and a physiological response and the ones that block the transmembrane proteins from interacting and thus preventing the intended physiological impact. In this study, we will examine the structures of the members of the class A G protein-coupled receptors (GPCRs) and the human glycoprotein Ecgp96 with the intention of designing drugs to interact with their respective binding sites.

Human cannabinoid CB1 receptor mediates hunger sensations

One example of a GPCR that has multiple active states is the human cannabinoid type 1 (CB1) receptor. CB1 is located primarily in the central nervous system and is so named because it is activated by the psychoactive component of cannabis, tetrahydrocannabinol, a chemical that causes feelings of hunger, euphoria, and relaxation. After ingestion of cannabis, people often report feeling excessively hungry. Therefore, marijuana is sometimes prescribed to AIDS or cancer patients who have lost their appetite and are having trouble maintaining a healthy weight. Marijuana has also makes the user feel content and relaxed. All of these psychoactive side effects—hunger, relaxation, and euphoria—are results induced through CB1. The human body produces its own endogenous cannabinoids, or endocannabinoids, which have the same psychoactive impact as cannabis, but are less efficacious (1). One type of endocannabinoid, anandamide—whose name is derived from the Sanskrit word for 'bliss'—is naturally found in chocolate (2). Cannabis, however, contains much larger doses of a more powerful chemical than that which the body normally produces (or what is found in chocolate), which is why it is impossible to get 'high' off of chocolate as one would from cannabis.

CB1 receptors are located throughout the brain including a region called the hypothalamus, which is responsible for appetite regulation amongst other things. In a normally functioning brain, this region receives neurotransmitters with messages telling us that we are full and don't need to eat. By blocking the release of these chemicals, CB1 ensures that the brain does not get those messages, resulting in hunger (3; 4).

Besides anecdotal evidence, many studies demonstrate the link between the endocannabinoid system and appetite. In an early experiment from 1971, thirteen people were separated into two groups. One group smoked two marijuana cigarettes and the other acted as the control by not smoking anything. After the marijuana ingestion, each person ate on average 46 marshmallows, whereas the people who didn't smoke anything ate on average only four marshmallows (5). In more recent studies, after given agonist drugs like cannabis that activate CB1, rats have been shown to work harder to obtain food, but when they are given drugs that block it, they do not try as rigorously (6). Mice genetically engineered to lack this receptor continue to receive hunger signals. They feel full sooner, so they do not work as hard to get food, and they do not respond to incentives (7). They also naturally eat fewer calories, which means that they weigh less and have less fat tissue (8). Research also shows that the addition of cannabinoid activators increases the nerve responses to sweeteners, which may explain how these molecules cause the improved palatability and increased motivation for food (6; 9).

Recently, scientists started trying to attack obesity by designing drugs to manipulate the role of CB1, in order to decrease food cravings and eliminate unwanted appetite. Studies show that obese rats have elevated levels of agonist endocannabinoids and CB1 receptors compared to lean ones (10). Agonists are molecules that cause a certain physiological effect—in this case for the user to think that they are hungry and need food. Blocker drugs fill the CB1's reaction site to prevent it from interacting with endocannabinoids and functioning normally. The cannabinoid receptor works by preventing the release of neurotransmitters that tell the brain we are full. Antagonists prevent this receptor from operating, so the neurotransmitters are released and the brain gets the "full" message sooner and more often. People receiving such drugs have a reduced appetite and, as a result, eat less and are more likely to lose weight.

In 1994, researchers at the pharmaceutical company Sanofi-Aventis understood that hunger and psychology are intricately connected and designed the antagonist/inverse agonist rimonabant to help people who practice healthy habits, but still need to lose weight (11). Rimonabant was supposed to be the next blockbuster drug. In a European study, patients with an average weight of 222 lbs taking the 20 mg dose lost 14.5 lbs over a year (12). This is an impressive number compared to the average 7.7 lbs lost over 2-3 years due to lifestyle changes alone (13). The drug, however, had serious side effects such as depression and increased risk of suicide. One study concluded that people participating in clinical trials were 2.4 times more likely to drop out due to mood disorders than those taking the placebo (14). It is unknown if these mood disorders are the

result of off-target interactions. If the structure of CB1 was known, a drug could be designed specifically for the CB1 binding site, thus limiting the possibility of it interacting with other proteins. However, CB1, like many other GPCRs, has never been crystallized, so its structure remains unknown.

G protein-coupled receptor (GPCR) activation procedure

GPCRs are some of the most popular drug targets with approximately 40% of prescription drugs designed to interact with GPCRs (15). This class of proteins is characterized by the seven alpha-helices that span the hydrophobic lipid bilayer. In its resting state, the GPCR resembles a funnel extending through the membrane, that is, the extracellular end of the helix bundle is wider than the intracellular end. When an activating ligand, or agonist, binds to a GPCR, it causes the receptor to change shape in a process called activation (**Figure 1.1**), in which the intracellular end of the receptor widens so that the global conformation resembles a hollow cylinder rather than a funnel. These conformational changes impact the receptor's interactions with other intracellular proteins, including the guanyl nucleotide-binding protein (G protein), initiating a signaling cascade and a physiological response (**Figure 1.2**). Agonists can be external substances like drugs, or they can be endogenous like neurotransmitters or hormones. They stabilize the active state, whereas inverse agonists, ligands that have the opposite physiological effect of agonists, stabilize the inactive state. Antagonists, ligands that prevent the agonists and inverse agonists from accessing the binding site, have no physiological impact. Like the inverse agonists, they will stabilize to the inactive state.

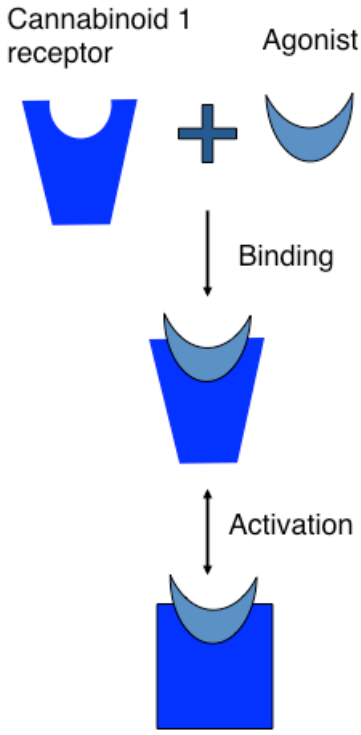


Figure 1.1 Schematic illustrations of the cannabinoid 1 (CB1) receptor undergoing activation after interaction with an agonist. CB1 (blue trapezoid) binds to agonists (light blue crescent), such as endocannabinoids or cannabis. This event causes the protein to undergo activation by changing its shape. The intracellular end of the receptor expands, so the whole complex resembles a cylinder. The receptor can now interact with proteins inside the cell.

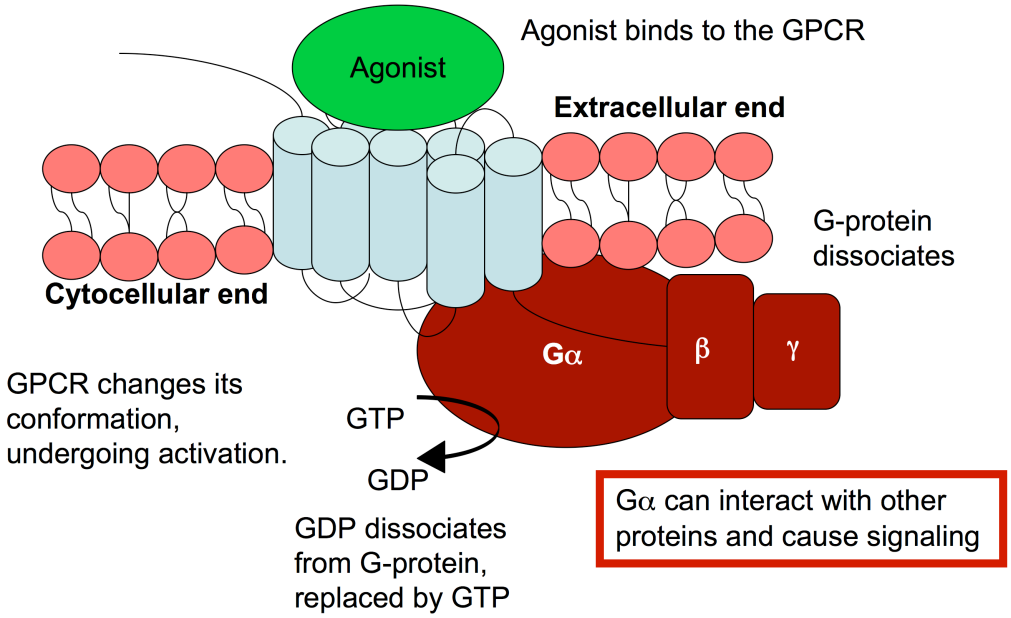


Figure 1.2 Schematic illustration explaining GPCR activation. Activation of GPCRs affects the body's behavior and by understanding the receptor's conformational changes, scientists can understand how to manipulate the subsequent physiological responses.

In the activated state, the GPCR's widened intracellular end has enough space for the heterotrimeric G protein to insert itself. When the G protein itself undergoes activation, guanosine diphosophate (GDP) is replaced with guanosine triphosophate (GTP), which causes the G protein to dissociate into α and βγ subunits that interact with other proteins leading to a physiological response. Multiple types of GPCRs, including the ones examined herein—the cannabinoid and opioid receptors—primarily interact with the inhibitory G_i protein (16-18). This protein prevents the production of the second messenger cyclic adenosine monophosphate (cAMP) as opposed to the G_s protein that stimulates its production (19).

Originally, it was believed that GPCRs switched between two conformations—an 'off' or inactive conformation and 'on' or active one. There is, however, experimental evidence that proves that GPCRs are dynamic and can exist in an ensemble of active or inactive states (20-23). At room temperature, only the lowest energy, most favorable, states are accessible. So when crystallized, the apo-receptor lacking a ligand will be captured in its inactive state. By adding an agonist to the receptor, the inactive state is destabilized and the activated G protein-coupled state becomes more energetically favorable. Sometimes a mutation can destabilize the inactive state and stabilize the active state without the assistance of an agonist or the G protein (24). These

mutants are referred to as 'constitutively active mutants,' or CAMs, and there are multiple examples of CAMs of GPCRs including the CB1 (25-28) and human mu-opioid (hMOR) receptors (29).

It is unknown how or what initiates the GPCR activation process. There are two main theories: conformational selection and induced fit (**Figure 1.3**). In conformational selection, the GPCR samples many different structures, some of them corresponding to the active ones, and an agonist binds to the receptor capturing the activated state. In induced fit, the agonist binds to the receptor in its inactive state, and the binding action induces activation and the conformational changes (30).

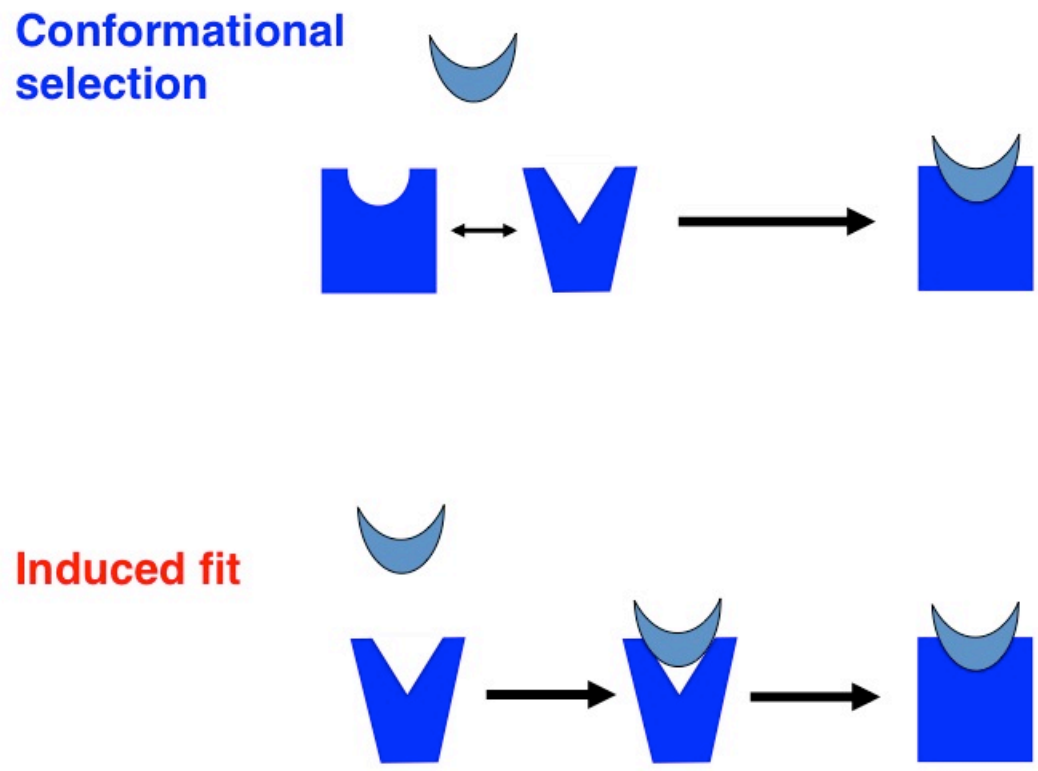


Figure 1.3 Illustration of the proposed receptor-ligand binding mechanisms, conformational selection and induced fit.

The transmembrane proteins' structures, however, are mostly unknown due to their instability outside the membrane leading to difficulty in isolation and crystallization. Thus, it is challenging to design drugs that interact with these receptors due to the lack of knowledge about their conformations and the drug's binding site. Even so, crystallographically characterized receptors

are just a single snapshot of a dynamic macromolecule. The GPCRs' activation mechanism remains a mystery, and the process of conformational changes going from an inactive conformation to an activated one is not well understood.

Due to the putative stability of the inactive state, most GPCR crystal structures correspond to this form including, but not limited to class-A GPCRs bovine rhodopsin (bRho) (31-41), squid (*Todarodes pacificus*) rhodopsin (sRho) (42; 43), human β_2 adrenergic receptor (inact$H\beta_2AR$) (44-46), turkey (*Meleagris gallopavo*) β_1 adrenergic receptor (inact$T\beta_1AR$) (47; 48), human A_{2A} adenosine receptor (inact$HA_{2A}AR$) (49; 50), human D3 dopamine receptor (hDD3R) (51), human CXCR4 receptor (hCXCR4) (52), human chemokine CCR5 receptor (hCCR5) (53), human sphingosine 1-phosphate 1 receptor (hS1P1) (54), mouse (*Mus musculus*) μ-opioid receptor (mMOR) (55), mouse δ-opioid receptor (mDOR) (56), human κ-opioid receptor (hKOR) (57), human NOP opioid receptor (hXOR) (58), human muscarinic M2 receptor (inactHMM2R) (59), and human muscarinic M3 receptor (hMM3R) (60). However, several crystal structures have captured one or more of the "active" forms of GPCRs: bovine opsin (bOps) (61; 62), bovine metarhodopsin II (meta II) (63), human β_2 adrenergic receptor (act$H\beta_2AR$) (64; 65), human A_{2A} adenosine receptor (act$HA_{2A}AR$) (66-68), human muscarinic M2 receptor (actHMM2R) (69), human neurotensin NTS1 receptor (70), and human P2Y12 receptor (71). These latter structures provide molecular level hints concerning the mechanisms of GPCR activation, but little is known about the structural transformations underlying the conversion of a GPCR from an inactive form through a constitutively active form to an even more constitutively active form. Only one GPCR has been crystallized in conjunction with the G_s protein (65), and the G_i protein has never been crystallized with a GPCR.

STATEMENT OF PROBLEM

While recent technological advances have caused the explosion in the number of crystallized GPCRs, there are still many receptors such as CB1 and CB2 that have not been crystallized and even more that have never been crystallized in the activated state. Even then, the crystallized conformation is just one of an ensemble of possible structures. GPCRs bind to a variety of intracellular proteins and ligands, and it is unknown how the GPCR structure changes to accommodate these new interactions.

Almost half the medication on the market is created to target a GPCR, but most of these drugs are designed without prior structural knowledge of the receptor, thus, increasing the likelihood of off-target side effects. A prime example of this problem is the CB1 blocker rimonabant. However, with predicted binding sites, drugs can be designed to interact specifically with these receptors.

PURPOSE OF STUDY

Here, we use computational techniques to predict an ensemble of structures for the GPCRs and their respective mutants representing various functional states. The Monte Carlo GEnSeMBLE (GPCR Ensemble of Structures in Membrane BiLayer Environment) method quickly and efficiently does a complete sampling of approximately a trillion GPCR structures to identify a small number of possible conformations. To validate the predicted structures, we dock known ligands to them in order to explain experimental site-directed mutagenesis and structure-activity relationship data. Based on the predicted binding sites, we design novel ligands to interact with the receptors. Collaborators will experimentally test these predictions. Ideally, we would calculate energies using quantum mechanical methods, but such computations are too expensive given the size of the receptors. Lastly, we use density functional theory to determine the reactivity of a high valent metal-oxo catalyst that could activate hydrocarbons.

OUTLINE OF THE THESIS

Chapter II: Methodology overview focusing on GPCR structural prediction methods and hierarchical ligand docking programs developed within the Goddard lab.

Chapter III: Structure prediction of the human cannabinoid type 1 (CB1) receptor and various mutants representing the inactive and constitutively active states.

Chapter IV: Docking of known inverse agonist/antagonist rimonabant to the ensemble of predict CB1 structures and design of new CB1-selective inverse agonist based on proposed binding site.

Chapter V: Molecular dynamics simulations of the predicted apo-CB1 receptor and the CB1 receptor-rimonabant complex to validate the stability of the structures and to see how they change over time.

Chapter VI: Structure prediction of the human cannabinoid type 2 (CB2) receptor and proposal of mutants that could capture the fully inactive state.

Chapter VII: Structure prediction of the human mu-opioid receptor (hMOR) and multiple mutants with various binding affinities for an opioid agonist as well as comparison to published crystal structure.

Chapter VIII: Docking of ligands to the human glycoprotein Ecgp96 with and without glycosylated units. Proposal of new ligands with greater predicted binding affinities and exploration of novel docking regions.

Chapter IX: Quantum mechanical calculations of the reactivity of the trans-dioxovanadium (V) porphyrin complex and free energy states involved in hydrocarbon activation.

REFERENCES

1. Nicoll RA, Alger BE (2004) The brain's own marijuana. Sci Am **1204**:68-75.
2. Di Tomaso E, Beltramo M, Piomelli D (1996) Brain cannabinoids in chocolate. Nature **382**:677-678.
3. Jo Y-H, Chen Y-J, Chua SC, Jr., Talmage DA, Role L (2005) Integration of endocannabinoid and leptin signaling in an appetite-related neural circuit. Neuron **48**:1055-1066.
4. Woods SC (2007) The endocannabinoid system: Mechanisms behind metabolic homeostasis and imbalance. Am J Med **120**:S9-S17.
5. Abel EL (1971) Effects of Marihuana on the Solution of Anagrams, Memory and Appetite. Nature **231**:260-261.
6. Kirkham TC, Rogers, Elizabeth K. (2010) Endocannabinoids in the aetiopathology of obesity – Central mechanisms. Drug Discov Today Dis Mech **7**:e163-e168.
7. Sanchis-Segura C, Cline BH, Marsicano G, Lutz B, Spanagel R (2004) Reduced sensitivity to reward in CB1 knockout mice. Psychopharmacology **176**:223-232.
8. Cota D, Marsicano G, Tschöp M, Grübler Y, Flachskamm C, Schubert M, Auer D, Yassouridis A, Thöne-Reineke C, Ortmann S, Tomassoni F, Cervino C, Nisoli E, Linthorst ACE, Pasquali R, Lutz B, Stalla GK, Pagotto U (2003) The endogenous cannabinoid system affects energy balance via central orexigenic drive and peripheral lipogenesis. J Clin Invest **112**:423-431.
9. Higgs S, Williams CM, Kirkham TC (2003) Cannabinoid influences on palatability: microstructural analysis of sucrose drinking after Delta(9)-tetrahydrocannabinol, anandamide, 2-arachidonoyl glycerol and SR141716. Psychopharmacology **165**:370-377.
10. Thanos PK, Ramalhete RC, Michaelides M, Piyis YK, Wang G-J, Volkow ND (2008) Leptin receptor deficiency is associated with upregulation of cannabinoid 1 receptors in limbic brain regions. Synapse **62**:637-642.
11. Rinaldi-Carmona M, Barth F, Heaulume M, Shire D, Calandra B, Congy C, Martinez S, Maruani J, Neliat G, Caput D, Ferrara P, Soubrie P, Breliere JC, Le Fur G (1994) SR1417

16A, a potent and selective antagonist of the brain cannabinoid receptor. FEBS Lett **350**:240-244.
12. Van Gaal LF, Rissanen AM, Scheen AJ, Ziegler O, Rossner S, RIO-Europe Study Group (2005) Effects of the cannabinoid-1 receptor blocker rimonabant on weight reduction and cardiovascular risk factors in overweight patients: 1-year experience from the RIO-Europe study. Lancet **365**:1389-1397.
13. Douketis JD, Macie C, Thabane L, Williamson DF (2005) Systematic review of long-term weight loss studies in obese adults: Clinical significance and applicability to clinical practice. Int J Obes Relat Metab Disord **29**:1153-1167.
14. Christensen R, Kristensen PK, Bartels EM, Bliddal H, Astrup A (2007) Efficacy and safety of the weight-loss drug rimonabant: A meta-analysis of randomised trials. Lancet **370**:1706-1713.
15. Filmore D (2004) It's a GPCR world. Mod Drug Discovery **7**:24-28.
16. Howlett AC, Fleming RM (1984) Cannabinoid inhibition of adenylate-cyclase. Pharmacology of the response in neuro-blastoma cell-membranes. Mol Pharmacol **26**:532-538.
17. Childers SR (1991) Opioid receptor-coupled second messenger systems. Life Sci **48**:1991-2003.
18. Hsia JA, Moss J, Hewlett EL, Vaughan M (1984) ADP-ribosylation of adenylate cyclase by pertussis toxin. Effects on inhibitory agonist binding. J Biol Chem **259**:1086-1090.
19. Hildebrandt JD, Sekura RD, Codina J, Iyengar R, Manclark CR, Birnbaumer L (1983) Stimulation and inhibition of adenylyl cyclases mediated by distinct regulatory proteins. Nature **302**:706-709.
20. Kenakin T (2002) Drug efficacy at G protein-coupled receptors. Annu Rev Pharmacol Toxicol **42**:349-379.
21. Kenakin T (2002) Efficacy at G-protein-coupled receptors. Nat Rev Drug Discov **1**:103-110.
22. Kenakin T (2003) Ligand-selective receptor conformations revisited: The promise and the problem. Trends Pharmacol Sci **24**:346-354.
23. Kenakin T, Miller LJ (2010) Seven transmembrane receptors as shapeshifting proteins: The impact of allosteric modulation and functional selectivity on new drug discovery. Pharmacol Rev **62**:265-304.
24. Abrol R, Kim S-K, Bray JK, Griffith AR, Goddard WA, III (2011) Characterizing and predicting the functional and conformational diversity of seven-transmembrane proteins. Methods **55**:405-414.
25. D'Antona A, Ahn KH, Kendall DA (2006) Mutations of CB_1 T210 produce active and inactive receptor forms: Correlations with ligand affinity, receptor stability, and cellular localization. Biochemistry **45**:5606-5617.
26. D'Antona A, Ahn KH, Wang L, Mierke DF, Lucas-Lenard J, Kendall DA (2006) A cannabinoid receptor 1 mutation proximal to the DRY motif results in constitutive activity and reveals intramolecular interactions involved in receptor activation. Brain Res **1108**:1-11.
27. Scott CE, Abrol R, Ahn KH, Kendall DA, Goddard WA, III (2013) Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations. Protein Sci **22**:101-113.
28. Ahn KH, Scott CE, Abrol R, Goddard WA, III, Kendall DA (2013) Computationally-predicted CB1 cannabinoid receptor mutants show distinct patterns of salt-bridges that correlate with their level of constitutive activity reflected in G protein coupling levels, thermal stability, and ligand binding. Proteins: Struct Funct Bioinform **81**:1304-1317.
29. Huang P, Visiers I, Weinstein H, Liu-Chen LY (2002) The local environment at the cytoplasmic end of TM6 of the mu opioid receptor differs from those of rhodopsin and monoamine receptors: Introduction of an ionic lock between the cytoplasmic ends of

helices 3 and 6 by a L6.30(275)E mutation inactivates the mu opioid receptor and reduces the constitutive activity of its T6.34(279)K mutant. Biochemistry **41**:11972-11980.
30. Deupi X, Kobilka BK (2010) Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. Physiology **25**:293-303.
31. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. Science **289**:739-745.
32. Standfuss J, Xie G, Edwards PC, Burghammer M, Oprian DD, Schertler GFX (2007) Crystal structure of a thermally stable rhodopsin mutant. J Mol Biol **372**:1179-1188.
33. Standfuss J, Edwards PC, D'Antona A, Fransen M, Xie G, Oprian DD, Schertler GFX (2011) The structural basis of agonist-induced activation in constitutively active rhodopsin. Nature **471**:656-660.
34. Stenkamp RE (2008) Alternative models for two crystal structures of bovine rhodopsin. Acta Crystallogr D Biol Crystallogr **64**:902-904.
35. Nakamichi H, Buss V, Okada T (2007) Photoisomerization mechanism of rhodopsin and 9-*cis*-rhodopsin revealed by x-ray crystallography. Biophys J **92**:L106-L108.
36. Salom D, Lodowski DT, Stenkamp RE, Le Trong I, Golczak M, Jastrzebska B, Harris T, Ballesteros JA, Palczewski K (2006) Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. Proc Natl Acad Sci USA **103**:16123-16128.
37. Okada T, Le Trong I, Fox BA, Behnke CA, Stenkamp RE, Palczewski K (2000) X-ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. J Struct Biol **130**:73-80.
38. Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y (2002) Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. Proc Natl Acad Sci USA **99**:5982-5987.
39. Okada T, Sugihara M, Bondar A-N, Elstner M, Entel P, Buss V (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. J Mol Biol **342**:571-583.
40. Teller DC, Okada T, Behnke CA, Palczewski K, Stenkamp RE (2001) Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). Biochemistry **40**:7761-7772.
41. Li J, Edwards PC, Burghammer M, Villa C, Schertler GFX (2004) Structure of bovine rhodopsin in a trigonal crystal form. J Mol Biol **343**:1409-1438.
42. Murakami M, Kouyama T (2008) Crystal structure of squid rhodopsin. Nature **453**:363-367.
43. Shimamura T, Hiraki K, Takahashi N, Hori T, Ago H, Masuda K, Takio K, Ishiguro M, Miyano M (2008) Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. J Biol Chem **283**:17753-17756.
44. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H-J, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. Science **318**:1258-1265.
45. Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola VP, Chien EYT, Velasquez J, Kuhn P, Stevens RC (2008) A specific cholesterol binding site is established by the 2.8 Å structure of the human β_2-adrenergic receptor. Structure **16**:897-905.
46. Rasmussen SGF, Choi H-J, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VRP, Sanishvili R, Fischetti RF, Schertler GFX, Weis WI, Kobilka BK (2007) Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. Nature **450**:383-387.
47. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG, Schertler GFX (2008) Structure of a β_1-adrenergic G-protein-coupled receptor. Nature **454**:486-491.

48. Moukhametzianov R, Warne T, Edwards PC, Serrano-Vega MJ, Leslie AGW, Tate CG, Schertler GFX (2011) Two distinct conformations of helix 6 observed in antagonist-bound structures of a β_1-adrenergic receptor. Proc Natl Acad Sci USA **108**:8228-8232.
49. Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, IJzerman AP, Stevens RC (2008) The 2.6 Å crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. Science **322**:1211-1217.
50. Doré AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, Hurrell E, Bennett K, Congreve M, Magnani F, Tate CG, Weir M, Marshall FH (2011) Structure of the adenosine A_{2A} receptor in complex with ZM241385 and the xanthines XAC and caffeine. Structure **19**:1283-1293.
51. Chien EYT, Liu W, Zhao QA, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, Stevens RC (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. Science **330**:1091-1095.
52. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science **330**:1066-1071.
53. Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Li J, Zhang W, Xie X, Yang H, Jiang H, Cherezov V, Liu H, Stevens RC, Zhao Q, Wu B (2013) Structure of the CCR5 Chemokine Receptor–HIV Entry Inhibitor Maraviroc Complex. Science **341**:1387-1390.
54. Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, Desale H, Clemons B, Cahalan SM, Schuerer SC, Sanna MG, Han GW, Kuhn P, Rosen H, Stevens RC (2012) Crystal structure of a lipid G protein-coupled receptor. Science **335**:851-855.
55. Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK, Granier S (2012) Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. Nature **485**:321-326.
56. Hurst D, Umejiego U, Lynch D, Seltzman H, Hyatt S, Roche M, McAllister S, Fleischer D, Kapur A, Abood M, Shi S, Jones J, Lewis D, Reggio P (2006) Biarylpyrazole inverse agonists at the cannabinoid CB1 receptor: Importance of the C-3 carboxamide oxygen/lysine3.28(192) interaction. J Med Chem **49**:5969-5987.
57. Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, Liu W, Thompson AA, Huang X-P, Carroll FI, Mascarella SW, Westkaemper RB, Mosier PD, Roth BL, Cherezov V, Stevens RC (2012) Structure of the human κ-opioid receptor in complex with JDTic. Nature **485**:327-332.
58. Barnett-Norris J, Hurst DP, Buehner K, Ballesteros JA, Guarnieri F, Reggio PH (2002) Agonist alkyl tail interaction with cannabinoid CB1 receptor V6.43/I6.46 groove induces a helix 6 active conformation. Int J Quant Chem **88**:76-86.
59. Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T, Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. Nature **482**:547-551.
60. Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO, Shaw DE, Weis WI, Wess J, Kobilka BK (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. Nature **482**:552-556.
61. Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP, Ernst OP (2008) Crystal structure of opsin in its G-protein-interacting conformation. Nature **455**:497-502.
62. Park JH, Scheerer P, Hofmann KP, Choe H-W, Ernst OP (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature **454**:183-187.
63. Choe HW, Kim YJ, Park JH, Morizumi T, Pai EF, Krauss N, Hofmann KP, Scheerer P, Ernst OP (2011) Crystal structure of metarhodopsin II. Nature **471**:651-655.

64. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SGF, Choi H-J, DeVree BT, Sunahara RK, Chae PS, Gellman SH, Dror RO, Shaw DE, Weis WI, Caffrey M, Gmeiner P, Kobilka BK (2011) Structure and function of an irreversible agonist-β_2 adrenoceptor complex. Nature **469**:236-240.
65. Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK (2011) Crystal structure of the β_2 adrenergic receptor-Gs protein complex. Nature **477**:549-555.
66. Xu F, Wu HX, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC (2011) Structure of an agonist-bound human A_{2A} adenosine receptor. Science **332**:322-327.
67. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG (2011) Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. Nature **474**:521-525.
68. Lebon G, Bennett K, Jazayeri A, Tate CG (2011) Thermostabilisation of an agonist-bound conformation of the human adenosine A_{2A} receptor. J Mol Biol **409**:298-310.
69. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Huber H, Pardon E, Valant C, Sexton PM, Christopoulos A, Felder CC, Gmeiner P, Steyaert J, Weis WI, Garcia C, Wess J, Kobilka BK (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. Nature **504**:101-106.
70. White JF, Noinaj N, Shibata Y, Love J, Kloss B, Xu F, Gvozdenovic-Jeremic J, Shah P, Shiloach J, Tate CG, Grisshammer R (2012) Structure of the agonist-bound neurotensin receptor. Nature **490**:508-513.
71. Zhang J, Zhang K, Gao Z-G, Paoletta S, Zhang D, Han GW, Li T, Ma L, Zhang W, Muller CE, Yang H, Jiang H, Cherezov V, Katritch V, Jacobson KA, Stevens RC, Wu B, Zhao Q (2014) Agonist-bound structure of the human P2Y12 receptor. Nature **509**:119-122.

Chapter II

METHODOLOGY: STRUCTURE AND BINDING SITE PREDICTION BASED ON COMPLETE SAMPLING MONTE CARLO

Adapted with permission from:
Scott CE, Abrol R, Ahn KH, Kendall DA, Goddard WA, III (2013) Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations. Protein Sci **22**:101-113.

Ahn KH, Scott CE, Abrol R, Goddard WA, III, Kendall DA (2013) Computationally-predicted CB1 cannabinoid receptor mutants show distinct patterns of salt-bridges that correlate with their level of constitutive activity reflected in G protein coupling levels, thermal stability, and ligand binding. Proteins: Struct, Funct, Bioinf **81**:1304-1317.

Bray JK, Abrol R, Goddard WA, III, Trzaskowski B, Scott CE (2014) SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations. Proc Natl Acad Sci U S A **111**:E72-E78.

The following chapter describes the background and technical details regarding the general methodology. Exceptions and further details for each computation will be noted in the respective chapters.

I. GEnSeMBLE procedure for generating an ensemble of receptor structures:

Most receptor structural prediction programs rely on homology modeling where the GPCR crystal structure's sequence is modified to match that of the target sequence. There are three problems with this procedure. First, while this method is particularly effective for predicting the structures of GPCRs closely related (over 40% sequence identity) to the crystal template, it is not useful for proteins that are not related to any crystallized GPCRs. For example, when we started predicting the structure for the CB1 receptor, it had a sequence identity to the crystallized GPCRs ranged from 9.96 to 16.10%, which means that none of the templates would be reliable since the sequences have so little in common (for more details see *Chapter III*). Second, many experimental and computational studies shown that single point mutations can cause significant changes in the binding site and hence receptor function. This can result from very modest changes in the interhelical interactions, which can cause rotations of the helix that completely modify the binding site. Homology methods do not address this issue. Homology modeling assumes that the changes in the sequence do not have a major impact on the structure, so they are not sensitive enough to capture the structural effect of a single change in the sequence. Third, homology modeling offers a limited number of predictions for the receptor structure. For example, to date, there are only six GPCRs crystallized in the active state (1-11)—only one crystallized in complex with the G_s protein (1). However, GPCRs bind to multiple agonists and intracellular proteins, so they have multiple activated (and possibly inactive) states (12-14) that homology modeling cannot capture.

Our lab has been developing methods (15-19) for predicting receptors' structures that address the three issues discussed above. The GEnSeMBLE (GPCR Ensemble of Structures in Membrane Bilayer Environment) procedure is based on first-principles, that is, we seek to predict the three-dimensional (3D) structure of the receptor based on its amino acid sequence with limited reliance on structural information from previously crystallized proteins. Briefly, we construct the seven individual helices from canonical helices using a combination of molecular dynamics (MD) simulations and minimization followed by aligning the helices to a GPCR template and determining the energies of trillions of receptor conformations by rotating and tilting the helices. This method allows us to thoroughly sample and calculate the energies of many receptor

conformations in a fast and efficient manner. Rather than obtaining a single solution for each GPCR, there is an ensemble of low energy conformations, and by constructing the helices from scratch, we build different conformations based on different receptor sequences, even if the only change is a single mutation. While these methods have been tested and validated with class A GPCRs, the procedures were designed to be general enough to apply to cases with little structural knowledge and no related crystallized templates. Currently, our co-workers are using these techniques to predict the structures of the distantly related class B GPCRs (20).

The procedure involves four steps outlined in **Figure 2.1**: 1) predicting the transmembrane (TM) helix regions of the receptor, 2) generating the template of the receptor bundle, 3) sampling a complete set of seven helix bundle conformations, and 4) identifying the lowest energy receptor conformations. We have verified our procedure using the few crystallized GPCR structures, and we find that our predicted structures are accurate (15; 17). Previously, Abrol *et al.* predicted the inactive human adenosine A_{2A} receptor (inact$HA_{A2}AR$) conformation starting from the inactive human β_2 adrenergic receptor (inact$H\beta_2AR$) template using the GEnSeMBLE method. After sampling, the inact$HA_{A2}AR$ predicted structure has a small 1.4 Å backbone root-mean-squared deviation (RMSD) with respect to the actual inact$HA_{A2}AR$ crystal (17).

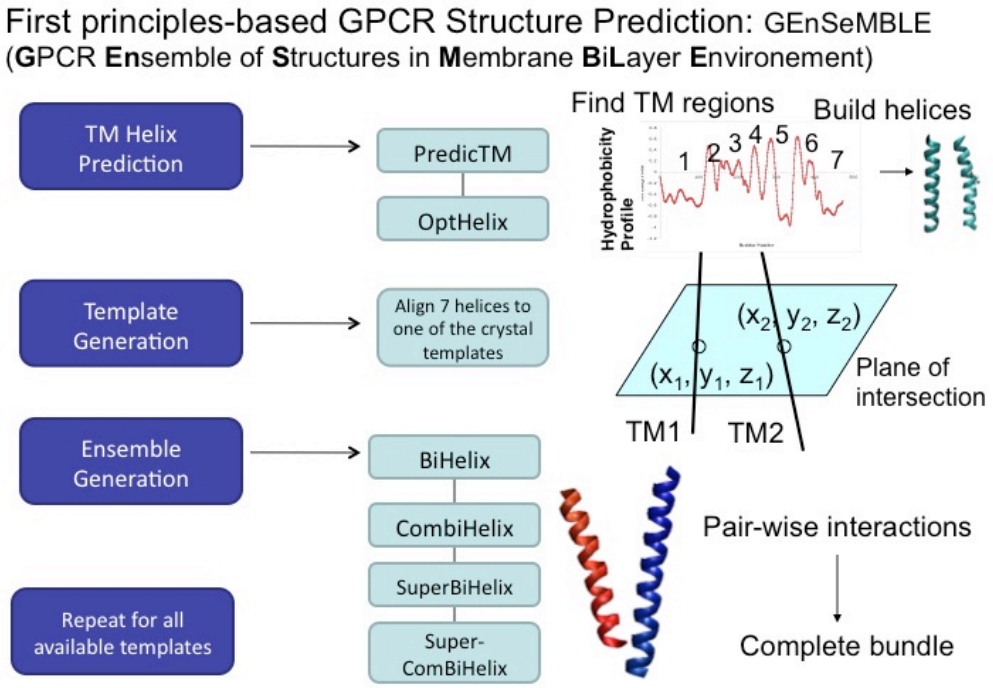


Figure 2.1 Outline of GEnSeMBLE method (15; 16; 18).

Part A: Predicting the wild-type (WT) receptor structures

Step 1: Predict transmembrane (TM) helix regions of the receptor

Based on the hydrophobic character of the target receptor amino acid sequence, we predict the TM regions using the PredicTM method. In the first step of PredicTM, a BLAST (21) search with an E-value of 0.1 is performed, or the program searches the BLAST database to find all related sequences based on our target receptor. The E-value represents the probability that a particular sequence pulled from the database is related to the target sequence. For example, a larger E-value means that a larger number of less related sequences will be found in the BLAST search. A smaller E-value means that a smaller number of sequences a higher conserved sequence identity will be found. With an E-value of 0.1, we find related seven helix GPCR sequences with sequence identities ranging in from 100% to single digits.

Our second step in PredicTM uses MAFFT software (22) based on the 'E-INS-i' method, which performs a multiple sequence alignment of the sequences identified by the BLAST search. MAFFT aligns the sequences according to their conserved portions and ignores the parts of the receptor with few conserved residues. This method is especially useful for GPCRs since the TMs are highly conserved, and the loop regions are not. This step is followed by using the Wimley-White octanol hydrophobicity scale (23) to assign a numeric hydrophobicity value to each amino acid position in the aligned sequence. The values are averaged over the whole set of aligned amino acids from each sequence and then these raw averages are averaged over traveling windows ranging in size from local sampling (7, 9, or 11 residues) to larger sampling (21 residues) to capture the hydrophobic character of the helix. The seven residue window corresponds to roughly two helix turns, while 21 residues is about the length of the full TM helix. Next, all eight traveling windows are averaged, so that each amino acid is assigned a single hydrophobicity value. These values for the target sequence are plotted to identify the seven obvious hydrophobic regions lying above the 0 baseline which correspond to the seven TM helical domains spanning the hydrophobic lipid bilayer providing the length and the hydrophobic center for each TM domain. The midpoint of the raw hydrophobic region or the residue bisecting the hydrophobic area for each helix provide the buoyant centers of the TM domains, with all seven residues lying in the same plane at the center of the membrane.

This hydrophobic analysis provides the raw hydrophobic regions of the helices, but GPCR crystal structures of bovine rhodopsin (PDB ID: 1U19) (24) and human chemokine receptor CXCR4

(hCXCR4) (PDB ID: 3ODU) (25) show that the hydrophobic helices can extend one or two turns out of the lipid bilayer into the solvent. To extend the hydrophobic regions of the helices, we developed a consensus helix capping method using three secondary structure prediction servers: Porter Protein Secondary Structure Prediction Server (26; 27), APSSP2: Advanced Protein Secondary Structure Prediction Server (28; 29), and PSIPRED: Protein Prediction Server (30-32).

Step 2: Optimize helix shapes for the seven TM segments

Step 2.1: Optimizing helices using molecular dynamics and minimization with OptHelix

Then, the OptHelix method was used to build and optimize the 3D structures of the TM domains. OptHelix performs a series of MD and energy minimizations to find the most stable structure of each individual helix. First, a canonical polyalanine helix is built matching the length of the predicted capped TM domains. Here, the glycine and proline residues from the original sequence are mutated into the helices using SCREAM (Side Chain Rotamer Excitation Analysis Method), a side chain optimization program (33). Proline and glycine residues are known helix-breakers that affect the backbone structure (34-36). The backbones of these TM domains are minimized to a root-mean squared (RMS) force of 0.5 kcal/mol/Å or to a maximum of 1000 steps in order to describe the modification on the backbone structure caused by mutating alanines to prolines or glycines. Previous studies have shown that the serine and threonine residues can either enhance or decrease the kink by interacting with the proline or amide groups within one helical turn (37). Thus, the appropriate alanine residues are changed to serine or threonine residues using SCREAM before doing a "warm-up" MD for each of the seven helices separately to prevent distortions that may be caused by the original assignment of velocities. The "warm-up" MD consists of heating in 50 K increments from 50 K to 250 K for 10 ps each with a 2 fs timestep for a total of 40 ps. Finally, the system is equilibrated with MD at 300 K for 2 ns with a 2 fs timestep. Ignoring the first 0.5 ns, snapshots are saved every 10 ps for a total of 150 snapshots. The snapshots with the lowest potential energy or with the RMSD closest to the average structure of the simulation are selected, and all residues are changed to their proper amino acid using SCREAM. The final structure is minimized to an RMS of 0.5 kcal/mol/Å or for a maximum number of 100 steps.

Step 2.2: Optimizing helices using homologized helices, or helix templates from crystallized GPCRs.

In addition to the above procedure with OptHelix, we also built helices using templates from crystal homology models. These templates bias the calculations to match the helix shape of the

crystallized GPCRs such as the inactive turkey β_1 adrenergic receptor, the inactive human β_2 adrenergic receptor, the inactive human adenosine A_{2A} receptor, and hCXCR4 to name a few. In each case we start with the crystal structure, remove the loops, mutate the residues to the target sequence, extend or shrink the length of the helices to match those of the OptHelix helices, and use the new helices to sample all 35.8 million combinations of the eta (rotation, η) angles using the same methodology described below. The helices are minimized 100 steps or to an RMS 0.0005 kcal/mol/Å. The homology helices are the same length as the OptHelix helices, so we can compare the total energy of the receptors since they all have the same number of atoms. Homology modeling is not expected to be effective for receptors distantly related to the target. Moreover, our results show dramatic changes in TM orientations with just a single point mutation, invalidating the fundamental assumption of homology modeling.

Step 3: Generate template for receptor bundle

The seven optimized helices are then aligned to experimental crystal templates. The initial GPCR crystal template provides initial values for the *x, y* positions and the theta (θ, tilt), phi (φ, sweep), and eta (η, rotation) angles for each helix. The *z*-position corresponds to the hydrophobic center determined from PredicTM in *Step 1* or the aligned homologized helix residue, with all seven hydrophobic centers on the same midpoint plane. Then, we exhaustively sample what we consider to be a complete set of θ, φ, and η angles to find the low energy receptor conformations.

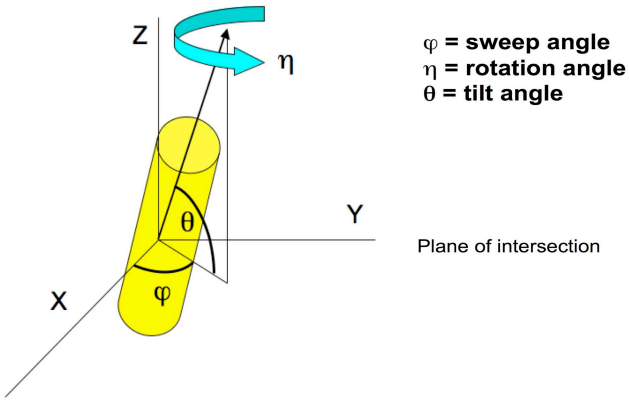


Figure 2.2 Six degrees of freedom for a helix. We can sample four of them (θ, φ, and η angles and *z*-axis) with the SuperBiHelix and SuperComBiHelix programs.

Step 4: Create ensemble of orientations of seven helices

We have found that it is essential to rotate all seven helices simultaneously to find the optimum combinations. Thus, considering 30° increments in η (which was found to be the minimum

sampling size), there are 12^7 = 35.8 million combinations, all of which must be examined in order to select the optimum packings of the bundles. The BiHelix and ComBiHelix programs were developed (15) to enable a complete sampling of all these 35.8 million conformations to identify the ensemble of low energy poses that might play a role in binding ligands and activation,.

The first step, BiHelix, independently considers just the 12 pairs of helices that are close enough to interact [1-2, 1-3, 1-7, 2-3, 2-4, 2-7, 3-4, 3-5, 3-6, 3-7, 4-5, 5-6, and 6-7]. For each of the 12^2=144 combinations of each pair, the side chains are optimized using SCREAM, and the energy is minimized for 10 steps. Summing the total intra- and interhelical values from these 12*144=1728 pairwise interaction energies leads to an energy estimate for all 35.8 million receptor bundles. Then, CombiHelix selects the combinations for the seven helix bundles predicted to have the lowest energy (38) and constructs the entire seven helix bundle to evaluate the total energy after using SCREAM to optimize the side chains and minimizing the energy for 10 steps. Ranking these bundles by their minimized total energy, we select the lowest energy bundle. This ComBiHelix bundle is chosen as the starting point for simultaneous optimization refinement of the tilt angles (θ and φ) while re-optimizing the η angles.

The lowest total energy ComBiHelix structure is used as the starting structure for SuperBiHelix, or further refinement in which the tilt angles (θ, φ) are optimized simultaneously with η. First, using SCREAM, the two terminal residues on each TM are replaced with alanine residues if they are one of the five charged residues (arginine, lysine, aspartic acid, glutamic acid, or histidine). This alanization of the terminal residues eliminates the strong electrostatic interactions at the ends of the helices, which may bias the calculations of the side chain interaction energies. This alanization step must be done when allowing variations in the θ and φ angles along with variations in the η angle. In SuperBiHelix (18), the η and φ angles are sampled over ±30° in 15° increments, while the θ is sampled over ±10° in 10°. With these ranges we consider $(5*5*3)^7$=13.3x10^{12} or ~13 trillion packings from which we select the top 2000 based on lowest total Super BiHelix combinations of pair-wise energies. These packings are built into full seven-helix bundles in SuperComBiHelix. Then, the full energy of each of these 2000 structures is evaluated after optimizing side chains and minimizing the energy.

Step 5: Choose the lowest energy structures

From the 2000 best receptor bundles from SuperComBiHelix, an ensemble of 10 low energy poses are chosen based on average energy rank. Here, we use four ways to calculate the energy.

Two ways use the normal charged form of aspartic acid, glutamic acid, lysine, arginine, and histidine, while the other two consider neutral forms for each of those residues. Two ways to combine the energies are considered. In addition to the total energy, we also consider only interhelical interactions with either charged or neutralized residues. From studies over many systems for which we can compare to x-ray crystal structures, we find that ranking the structures using all four methods and averaging this ranking leads to the most reliable predictions. The all atom Dreiding force field (39) is used for all energy calculations.

Part B: Predicting the mutant receptor structures

Sometimes it is necessary to build mutants of the WT receptors. Site-directed mutagenesis is useful for testing lingand binding affinity or functionality. We are especially interested in constitutively active mutations (CAMs) because they can significantly change the functionality of the receptor by changing a single residue. GEnSeMBLE is a useful method for structure prediction because we construct the mutants starting from the helices and observe the global impact a single residue can have on the entire receptor conformation.

Mutate the specified residue using SCREAM

To predict the ensemble of low energy mutant structures, we use the same methodology as described for the WT receptor. Here, the WT helices built in Part A Step 1 are used to generate the receptor bundle, but one, two, or even three residues are mutated using SCREAM. The subsequent steps (*Steps 2 - 5*) are the same for the mutants as they are for the WT receptor (see **Part A**), where full helix bundle conformational sampling is performed for each mutant.

II. DarwinDock and GenDock procedure for completely sampling ligand poses in the binding site

These predicted conformations from Part I can be further validated by docking known ligands to them and explaining experimental site-directed mutagenesis and structure-activity relationship (SAR) data. Site-directed mutagenesis data is useful because it indicates which residues are potentially important for anchoring the ligand in the binding site. If the ligand's binding affinity is significantly decreased when a residue is mutated, either that residue has strong interactions with the ligand or the mutation caused the receptor to undergo a conformational change. Usually, the assumption is that the loss in binding affinity is due to the former, but in this work, we check the possibility of the latter. SAR data indicates which portion of the ligand is important for

interactions with the receptor. A particular portion of the ligand is modified multiple times to generate many different ligands. The binding affinity is measured for these ligands and large changes in affinity indicate which parts of the ligand are important for interacting with the binding site. The objective is to explain these experimental data on the molecular level with our computational studies. For example, it is expected that the ligand will have strong energetic interactions with residues that have a large impact on experimental binding affinity. Based on our predicted binding site, we also expect to be able to explain why modifications in the ligand structure impacts the interactions with the receptor and thus the binding affinity and to show trends between our calculated binding and cavity energies and the experimental binding affinities.

Many docking programs rely on training sets, that is, they are based on information of previously crystallized protein-ligand complexes. Since there is a dearth of information about GPCR structures, these training sets are not relevant to our systems. Our hierarchical docking methods, DarwinDock and GenDock (40-44), generate thousands of ligand conformations to thoroughly and completely sample the receptor's binding site before refining and enriching a small number of poses and determining the best one energetically. Like GEnSeMBLE, this procedure relies on completely sampling all potentially important conformations in an efficient manner. The receptor-ligand complexes are chosen using an approach that is not biased by other experimental data. First, we need to construct ligands to dock to various conformations of the receptors and determine which receptor conformation the ligand prefers to bind to.

Part A: Ligand preparation

The DarwinDock program docks rigid ligands to rigid receptors, so it is necessary to dock multiple diverse ligand conformations. For each ligand, we construct each structure with Maestro software (45) and preform a conformational search with MacroModel software (46). We use sytematic and extended torsional sampling where the selected rotatable bonds are rotated 360° in 30° increments. We save ligand conformations that fall within an energy window of 10 kJ/mol and an RMSD window of 0.5 Å. The conformational search is conducted with the OPLS 2005 force field (47) and in a dielectric of 80.37 to match water. We preform two rounds of clustering for each ligand—the first round we clustered ligands with a 2.0 Å diversity followed by another round with a 1.0 Å. We calculate the Mulliken populations of each atom with Jaguar software (48) using Density Functional Theory (DFT) with the B3LYP functional and the 6-31G** basis set. Sometimes one of the conformations for the ligand is constructed from existing coordinates from a crystallized small molecule deposited in the Cambridge Structural Database

(49). Then each ligand is minimized with SGB solvation (50) for 100 steps or to a convergence threshold RMS of 0.2 kcal/mol/Å with MPSim software (38).

Part B: Docking procedures

Using the programs DarwinDock and GenDock (40-44) developed within our lab, we dock multiple ligand conformations to two or three selected regions that encompass the binding site of the low energy structures of each of the receptors. The docking procedure consists of two steps, DarwinDock (**Figure 2.3**), which generates enough ligand poses to completely sample the binding site, and GenDock (**Figure 2.4**), which enriches the lowest energy complexes from the DarwinDock step. **Figures 2.3** and **2.4** show a general overview of the methods, and technical details are summarized below.

Step 1: Generate ligand-receptor complexes with DarwinDock

DarwinDock (**Figure 2.3**) combines the completeness of generating poses to thoroughly sample the space of the binding pocket, indicated by the sphere regions, and the clustering of similar poses together to reduce the computational cost of energy scoring by only scoring diverse candidates. In preparation for this step, the receptors are 'alanized' by replacing the bulky, hydrophobic residues (phenylalanine, isoleucine, leucine, methionine, tyrosine, valine, tryptophan) with alanine residues. This residue substitution increases the amount of sampling space in the binding site because the larger residues will not clash with the ligand. We believe that the stronger interactions occur with the polar ones, so the ligands will be anchored in the binding site by hydrogen bonds and salt bridges, while the phobic residues can fill in around these previously formed contacts.

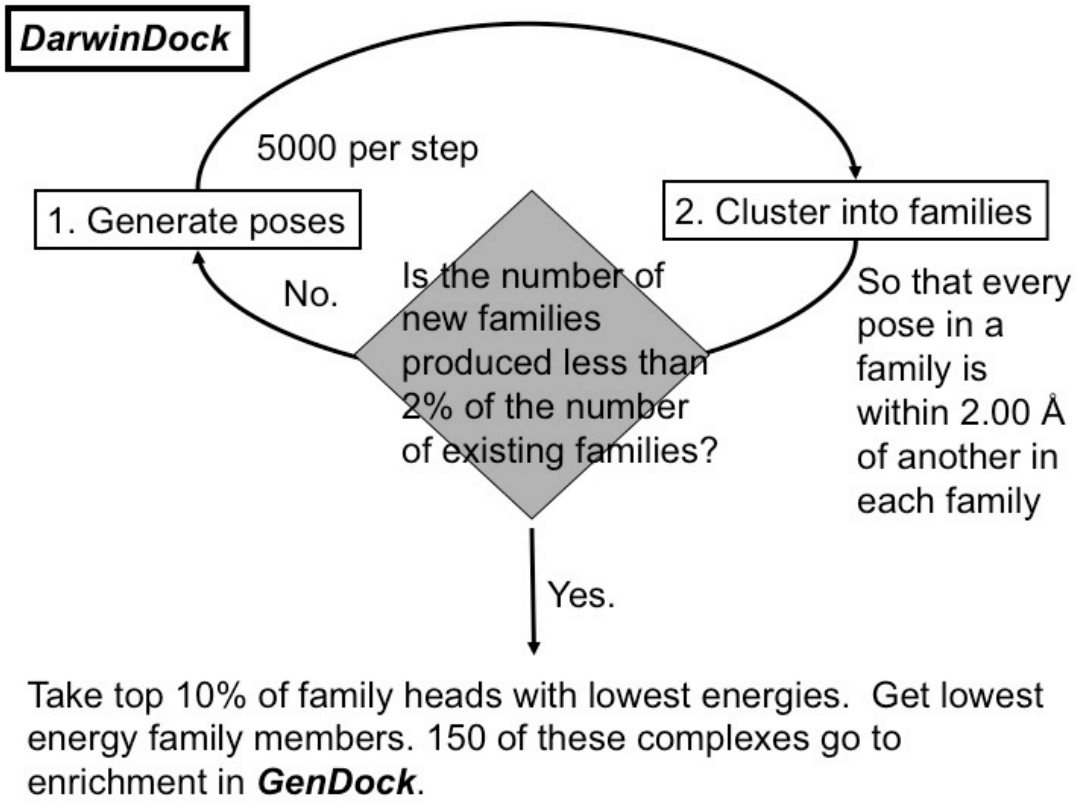


Figure 2.3 DarwinDock procedure diagram. The above schematic explains the process DarwinDock uses to generate approximately 150 complex poses that will enter GenDock, the enrichment phase.

The first step is to use Dock6 (51) to generate thousands of ligand poses in the alanized receptor. Each iteration of DarwinDock consisted of a series of steps: first, Dock6 produced 5000 ligand-receptor complex poses and then the ligand poses were grouped into 'families' using Vornoi clustering where every member of the family was within a 2.00 Å RMSD of another. A ligand pose is discarded if it clashes or bumps the receptor residues more than a specified number of times depending on the size of the ligand. These steps are repeated until 'completeness' is reached. The definition of 'completeness' is when the number of new families created is less than 2% of the number of existing families indicating that the binding site has been thoroughly and completely sampled. A representative pose termed the 'family head' is chosen for each family. Each family head's energy is calculated, and 10% of the family heads with the lowest energies are chosen. The program then scores the energies of all the poses in the respective families of the identified low energy family heads. One hundred and fifty receptor-ligand poses are selected for the next step, GenDock, based on their energies. Fifty poses are chosen according to the lowest phobic energy, another 50 poses according to the lowest polar energy, and 50 poses according to the lowest total energy.

Step 2: Refine ligand-receptor complexes with GenDock

The GenDock program used three modules to refine and enrich the 150 docked ligand-receptor poses generated by DarwinDock. **Figure 2.4** outlines this procedure using the WT CB1 receptor and the cannabinoid agonist WIN55212-2 as examples. The first module uses the SCREAM program (33) to replace alanines with the original hydrophobic residues (refer to the DarwinDock section) and to sample the side chain rotamers of any residue within 4.00 Å of any ligand pose. Then the entire complex including the receptor backbone is minimized for 10 steps. The second module neutralizes the receptor by adding protons to the acidic residues (aspartic acid, glutamic acid) and removing a proton from the basic residues (lysine, arginine, and histidine). Depending upon the pKa of the ligand, it will either gain or lose protons until its total charge is zero. The 10 steps of minimization are repeated for the entire complex. These neutralized energies are more realistic for comparisons between different complexes (52). The last module consists of a full relaxation of the entire complex with 50 steps of minimization with the Dreiding III FF (39). We believe that this FF is more reliable than the energy scoring function used by Dock6.

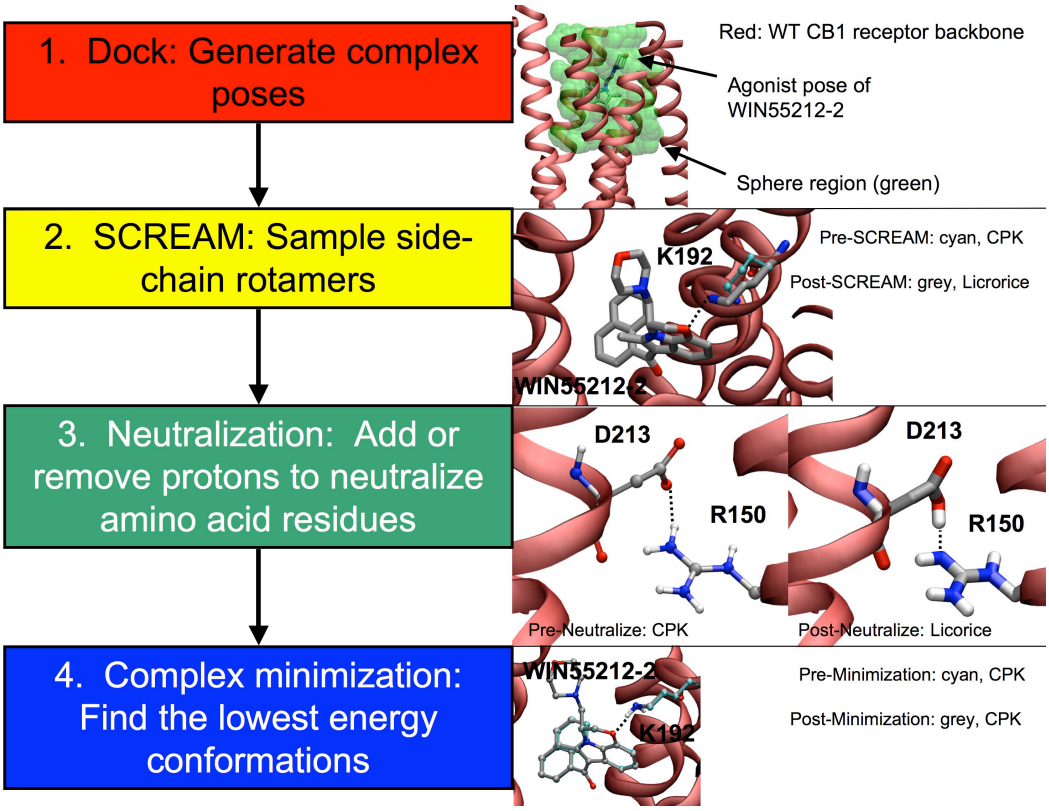


Figure 2.4 GenDock procedure diagram. Lowest energy poses from DarwinDock procedure undergo enrichment phase illustrated above.

Step 3: Select the best complex energetically

We modified the Dreiding III FF by replacing the exponential-6 term with the Lennard-Jones (LJ) one to better account for repulsive actions between atoms. Using MPSim software (38), we minimized the selected complexes from the previous step with the modified Dreiding III-LJ force field in vacuum for 50 steps or to an RMS threshold of 0.5 kcal/mol/Å.

There are multiple types of energies that can accurately identify the correct ligand pose in the receptor's binding site. The binding energy (**SnapBE**) is defined as the energy difference between the complex and the sum of the receptor and ligand energies. Most of the docking complexes herein are ordered according to the binding energy with the ligand strain and ligand solvation added (**SnapBE-strain-ligsolv**). The cavity energy (**UCav**) is the energy between the ligand and the residues in the unified binding site, or the list of residues interacting with each ligand pose is combined to create a unified list. The energy between the ligand and all the residues in this list is calculated.

III. Applications to GPCRs and other membrane proteins

The focus of this thesis is how we applied these structural prediction methods to GPCRs and another membrane protein, the human glycoprotein Ecgp96, in order to obtain accurate predictions and understand how these proteins interact with small molecules and drugs. The human CB1 receptor is a stellar example of the application of these methods to predict the structure of a TM receptor that has not been crystallized. At the start of this project, none of the crystallized GPCRs were closely related to CB1. Thus, we used the GEnSeMBLE method to perform a thorough sampling of trillions of CB1 conformations to obtain an ensemble of low energy structures. We repeated this procedure for two single-point mutants that had significantly different levels of activity compared to the WT. The resulting structures for the three receptors are dramatically different on a global scale and had different patterns of charged interactions on the intracellular end where the G protein binds. Most importantly, the structural differences explain the experimentally confirmed functional differences. To test our structural understanding, we designed double and triple mutants with specific levels of activity in mind. Subsequent experiments later confirmed our predictions proving that certain charged interactions are important for stabilizing the active or inactive conformations of the CB1 receptor. Hence, the above methodology is sensitive enough to detect the impact that a single residue has on the global conformation of the GPCR, which cannot be done with the homology methods.

To further support our predicted structures, we docked the known CB1 inverse agonist/antagonist rimonabant to our ensemble of CB1 predicted structures and were able to explain experimental site-directed mutagenesis and SAR data. Our proposed rimonabant binding site is different from that which has been published previously (53-57). Based on the newly predicted binding site, we identified ligands in the PubChem database (58) that are commercially available and have not been tested with the CB1 receptor. We predict that these ligands are strong CB1-selective inverse agonists, which could be effective anti-obesity drugs. Thus, these methods have been very successful in providing insight into the structure and function of CB1 as well as other GPCRs discussed herein.

REFERENCES

1. Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK (2011) Crystal structure of the β_2 adrenergic receptor-G_s protein complex. Nature **477**:549-555.
2. Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK (2011) Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. Nature **469**:175-180.
3. Choe HW, Kim YJ, Park JH, Morizumi T, Pai EF, Krauss N, Hofmann KP, Scheerer P, Ernst OP (2011) Crystal structure of metarhodopsin II. Nature **471**:651-655.
4. Park JH, Scheerer P, Hofmann KP, Choe H-W, Ernst OP (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature **454**:183-187.
5. Standfuss J, Edwards PC, D'Antona A, Fransen M, Xie G, Oprian DD, Schertler GFX (2011) The structural basis of agonist-induced activation in constitutively active rhodopsin. Nature **471**:656-660.
6. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG (2011) Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. Nature **474**:521-525.
7. Xu F, Wu HX, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC (2011) Structure of an agonist-bound human A_{2A} adenosine receptor. Science **332**:322-327.
8. Lebon G, Bennett K, Jazayeri A, Tate CG (2011) Thermostabilisation of an agonist-bound conformation of the human adenosine A_{2A} receptor. J Mol Biol **409**:298-310.
9. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Huber H, Pardon E, Valant C, Sexton PM, Christopoulos A, Felder CC, Gmeiner P, Steyaert J, Weis WI, Garcia C, Wess J, Kobilka BK (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. Nature **504**:101-106.
10. White JF, Noinaj N, Shibata Y, Love J, Kloss B, Xu F, Gvozdenovic-Jeremic J, Shah P, Shiloach J, Tate CG, Grisshammer R (2012) Structure of the agonist-bound neurotensin receptor. Nature **490**:508-513.
11. Zhang J, Zhang K, Gao Z-G, Paoletta S, Zhang D, Han GW, Li T, Ma L, Zhang W, Muller CE, Yang H, Jiang H, Cherezov V, Katritch V, Jacobson KA, Stevens RC, Wu B, Zhao Q (2014) Agonist-bound structure of the human P2Y12 receptor. Nature **509**:119-122.

12. Kobilka BK, Deupi X (2007) Conformational complexity of G-protein-coupled receptors. Trends Pharmacol Sci **28**:397-406.
13. Kenakin T, Miller LJ (2010) Seven transmembrane receptors as shapeshifting proteins: The impact of allosteric modulation and functional selectivity on new drug discovery. Pharmacol Rev **62**:265-304.
14. Deupi X, Kobilka BK (2010) Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. Physiology **25**:293-303.
15. Abrol R, Bray JK, Goddard WA, III (2012) BiHelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. Proteins: Struct, Funct, Bioinf **80**:505-518.
16. Abrol R, Griffith A, Bray J, Goddard WA, III. Structure prediction of G protein-coupled receptors and their ensemble of functionally important conformations. In: Vaidehi N, Klein-Seetharaman J, Eds. (2012) Membrane Protein Structure: Methods and Protocols. Humana Press, New York, NY.
17. Abrol R, Kim S-K, Bray JK, Griffith AR, Goddard WA, III (2011) Characterizing and predicting the functional and conformational diversity of seven-transmembrane proteins. Methods **55**:405-414.
18. Bray JK, Abrol R, Goddard WA, III, Trzaskowski B, Scott CE (2014) SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations. Proc Natl Acad Sci USA **111**:E72-E78.
19. Abrol R, Kim S-K, Bray JK, Trzaskowski B, Goddard WA, III (2013) Conformational ensemble view of G protein-coupled receptors and the effect of mutations and ligand binding. Methods Enzymol **520**:31-48.
20. Kirkpatrick A, Heo J, Abrol R, Goddard WA, III (2012) Predicted structure of agonist-bound glucagon-like peptide 1 receptor, a class B G protein-coupled receptor. Proc Natl Acad Sci USA **109**:19988-19993.
21. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25**:3389-3402.
22. Katoh K, Kuma K-I, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res **33**:511-518.
23. Wimley WC, Creamer TP, White SH (1996) Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry **35**:5109-5124.
24. Okada T, Sugihara M, Bondar A-N, Elstner M, Entel P, Buss V (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. J Mol Biol **342**:571-583.
25. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science **330**:1066-1071.
26. Pollastri G, McLysaght A (2004) Porter: A new, accurate server for protein secondary structure prediction. Bioinformatics **21**:1719-1720.
27. http://distill.ucd.ie/porter/.
28. Raghava GPS. APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. (2002) CASP5. pp. A-132.
29. http://www.imtech.res.in/raghava/apssp2/.
30. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol **292**:195-202.
31. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT (2005) Protein structure prediction servers at university college london. Nucleic Acids Res **33**:W36-W38.

32. http://bioinf.cs.ucl.ac.uk/psipred/: 2002-2008.
33. Kam VWT, Goddard WA, III (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. J Chem Theory Comput **4**:2160-2169.
34. Cordes FS, Bright JN, Sansom MSP (2002) Proline-induced distortions of transmembrane helices. J Mol Biol **323**:951-960.
35. Yohannan S, Faham S, Yang D, Whitelegge JP, Bowie JU (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. Proc Natl Acad Sci USA **101**:959-963.
36. Chakrabarti P, Chakrabarti S (1998) C-H . . . O hydrogen bond involving proline residues in alpha-helices. J Mol Biol **284**:867-873.
37. Deupi X, Olivella M, Govaerts C, Ballesteros JA, Campillo M, Pardo L (2004) Ser and Thr residues modulate the conformation of Pro-kinked transmembrane α-helices. Biophys J **86**:105-115.
38. Lim KT, Brunett S, Iotov M, McClurg RB, Vaidehi N, Dasgupta S, Taylor S, Goddard WA, III (1997) Molecular dynamics for very large systems on massively parallel computers: The MPSim program. J Comput Chem **18**:501-521.
39. Mayo SL, Olafson BD, Goddard WA, III (1990) DREIDING: A generic force field for molecular simulations. J Phys Chem **94**:8897-8909.
40. Kim S-K, Li Y, Abrol R, Heo J, Goddard WA, III (2011) Predicted Structures and Dynamics for Agonists and Antagonists Bound to Serotonin 5-HT2B and 5-HT2C Receptors. J Chem Inf Model **51**:420-433.
41. Kim S-K, Riley L, Abrol R, Jacobson KA, Goddard WA, III (2011) Predicted structures of agonist and antagonist bound complexes of adenosine A(3) receptor. Proteins: Struct, Funct, Bioinf **79**:1878-1897.
42. Goddard WA, III, Kim S-K, Li Y, Trzaskowski B, Griffith AR, Abrol R (2010) Predicted 3D structures for adenosine receptors bound to ligands: Comparison to the crystal structure. J Struct Biol **170**:10-20.
43. Floriano WB, Vaidehi N, Zamanakos G, Goddard WA (2004) HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. J Med Chem **47**:56-71.
44. Nair N, Kudo W, Smith MA, Abrol R, Goddard WA, III, Reddy VP (2011) Novel purine-based fluoroaryl-1,2,3-triazoles as neuroprotecting agents: Synthesis, neuronal cell culture investigations, and CDK5 docking studies. Bioorg Med Chem Lett **21**:5649-5649.
45. Maestro, version 7.5. (2006). Schrodinger, LLC, New York, NY.
46. MacroModel, 9.5. (2007). Schrodinger, LLC, New York, NY.
47. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc **118**:11225-11236.
48. Jaguar, version 6.5. (2006). Schrodinger, LLC, New York, NY.
49. Allen FH (2002) The Cambridge Structural Database: A quarter of a million crystal structures and rising. Acta Crystallographica Section B-Structural Science **58**:380-388.
50. Ghosh A, Rapp CS, Friesner RA (1998) Generalized born model based on a surface integral formulation. J Phys Chem B **102**:10983-10990.
51. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol **161**:269-288.
52. Bray JK, Goddard WA, III (2008) The structure of human serotonin 2C G-protein-coupled receptor bound to agonists and antagonists. J Mol Graphics Modell **27**:66-81.
53. McAllister SD, Rizvi G, Anavi-Goffer S, Hurst DP, Barnett-Norris J, Lynch DL, Reggio PH, Abood ME (2003) An aromatic microdomain at the cannabinoid CB_1 receptor constitutes an agonist/inverse agonist binding region. J Med Chem **46**:5139-5152.
54. Hurst DP, Lynch DL, Barnett-Norris J, Hyatt SM, Seltzman HH, Zhong M, Song ZH, Nie JJ, Lewis D, Reggio PH (2002) N-(Piperidin-1-yl)-5-(4-chlorophenyl)-1-(2,4-

dichlorophenyl)-4-methyl-1H -pyrazole-3-carboxamide (SR141716A) interaction with LYS 3.28(192) is crucial for its inverse agonism at the cannabinoid CB1 receptor. Mol Pharmacol **62**:1274-1287.

55. Lange JHM, Kruse CG (2005) Medicinal chemistry strategies to CB1 cannabinoid receptor antagonists. Drug Discovery Today **10**:693-702.
56. Salo OMH, Lahtela-Kakkonen M, Gynther J, Jarvinen T, Poso A (2004) Development of a 3D model for the human cannabinoid CB1 receptor. J Med Chem **47**:3048-3057.
57. Silvestri R, Cascio MG, La Regina G, Piscitelli F, Lavecchia A, Brizzi A, Pasquini S, Botta M, Novellino E, Di Marzo V, Corelli F (2008) Synthesis, cannabinoid receptor affinity, and molecular modeling studies of substituted 1-aryl-5-(1H-pyrrol-1-yl)-1H-pyrazole-3-carboxamides. J Med Chem **51**:1560-1576.
58. Bolton E, Wang Y, Thiessen P, Bryant S. PubChem: Integrated platform of small molecules and biological activities. (2008) Annual Reports in Computational Chemistry Elsevier Ltd., Oxford, UK, pp. 217-240.

Chapter III

STRUCTURAL BASIS OF HUMAN CANNABINOID CB1 G PROTEIN-COUPLED RECEPTOR ACTIVATION

The GTPγS assay results shown herein were preformed by Dr. Kwang Hyun Ahn and Professor Debra Kendall.

Adapted with permission from:
Scott CE, Abrol R, Ahn KH, Kendall DA, Goddard WA, III (2013) Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations. Protein Sci **22**:101-113.

Ahn KH, Scott CE, Abrol R, Goddard WA, III, Kendall DA (2013) Computationally-predicted CB1 cannabinoid receptor mutants show distinct patterns of salt-bridges that correlate with their level of constitutive activity reflected in G protein coupling levels, thermal stability, and ligand binding. Proteins: Struct, Funct, Bioinf **81**:1304-1317.

Bray JK, Abrol R, Goddard WA, III, Trzaskowski B, Scott CE (2014) SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations. Proc Natl Acad Sci U S A **111**:E72-E78.

ABSTRACT

There is considerable interest in determining the activation mechanism of G protein-coupled receptors (GPCRs), one of the most important types of proteins for intercellular signaling. Recently it was demonstrated, for the cannabinoid CB1 GPCR, that a single mutation T210A could make CB1 completely inactive whereas T210I makes it essentially constitutively active. To obtain an understanding of this dramatic dependence of activity on mutation, we used first principles-based methods to predict the ensemble of low energy seven helix conformations for the wild-type (WT) and mutants (T210A, T210I, and L207A). We find that the transmembrane (TM) helix packings depend markedly on these mutations, leading for T210A to both TM3+TM6 and TM2+TM6 salt bridge couplings in the cytoplasmic face that explain the inactivity of this mutant. In contrast, T210I and L207A have no such couplings across the receptor explaining the ease in activating these mutants. The WT receptor only has the TM3+TM6 coupling, known to be broken upon GPCR activation. To test the hypothesis that the TM2+TM6 salt bridge ensures inactivity, we predicted double mutants and a triple mutant that would convert the inactive mutant to normal activity, and then our collaborators confirmed this experimentally. We also proposed a adding a second mutation to the constitutively active L207A receptor to make it fully inactive, which was also confirmed experimentally. This CB1 activation mechanism, or one similar to it, may be applicable to other constitutively active class A GPCRs as well.

INTRODUCTION

Activation of G protein-coupled receptors (GPCRs) by extracellular sensory and non-sensory signals affects multiple intracellular signaling cascades by modulating G protein and/or β arrestin-coupled pathways (1). This pleiotropic consequence on GPCR activation is facilitated by their remarkable conformational flexibility, enabling the formation of multiple distinct conformations that can influence function. The human cannabinoid type 1 (CB1) receptor, which is located mainly in the central nervous system, displays partial constitutive activity. We have shown previously (2) and confirmed using GTPγS assays (**Figure 3.1),** that mutating the threonine residue T210 near the cytoplasmic end of the third transmembrane region (TM3) to alanine makes CB1 inactive, whereas mutating it to isoleucine makes it highly constitutively active (2). The residue T210 will be referred to as T3.46 throughout the remainder of this thesis [using the Ballesteros-Weinstein numbering scheme (3)], indicating that it is located one helix turn above the conserved R3.50 residue on the intracellular end of TM3. Changes in binding

profiles for the inverse agonist rimonabant in the presence and absence of G protein-coupling indicated that the T3.46I receptor adopts the active state, while the T3.46A receptor is in an inactive state (2). We found experimentally that a high concentration of GTPγS precludes G protein-coupling, promoting the inactive form. Under these conditions, T3.46I, which otherwise has a weak binding affinity for inverse agonists, displayed a 30-fold enhancement in its affinity for rimonabant. In comparison, the wild-type (WT) receptor displayed a 5-fold increase in rimonabant affinity, while the T3.46A receptor showed a negligible increase in affinity for this ligand. This hierarchy of rimonabant affinity changes parallels the ease with which each receptor adopts an inactive conformation (2). These major activity changes upon single point mutations provide an excellent system for probing the molecular mechanisms of GPCR activation, allowing us to follow the structural changes from the fully inactive form, through the weakly constitutively active form typically seen in WT receptors, to the highly constitutively active form.

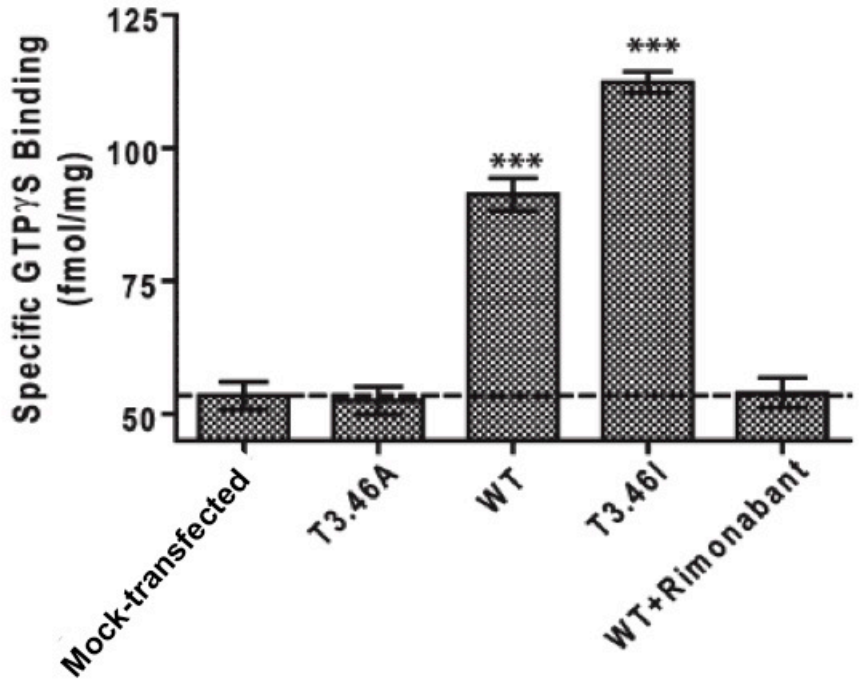


Figure 3.1 Comparison of basal GTPγS binding to HEK293 cell membranes expressing the CB1 receptors. The level of [^{35}S]GTPγS binding was measured in the absence of ligand for the WT, T3.46I, and T3.46A receptors. The [^{35}S]GTPγS binding for the mock-transfected, and for the membrane expressing the WT receptor with inverse agonist rimonabant treatment, are shown for comparison. Data are presented as specific binding of GTPγS to the membrane preparation. Nonspecific binding was determined in the presence of 10 μM unlabeled GTPγS. Each data point represents the mean ± S.E.M. of at least three independent experiments performed in duplicate. Statistical significance of the differences between the level of each receptor and that of a mock-transfected sample was assessed using one-way analysis of variance and Bonferroni's post-hoc test; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. The dashed line indicates the level of non CB1-mediated GTPγS binding obtained from [^{35}S]GTPγS binding to the mock-transfected membrane sample.

To evaluate the basal G protein-coupling activity of the WT, T3.46A, and T3.46I receptors, we measured the level of [^{35}S]GTPγS binding to HEK293 cell membranes expressing the receptors in the absence of ligand **(Figure 3.1)**. Specific GTPγS binding for the T3.46A receptor is 52.6 fmol/mg in the absence of the ligand, which is comparable to the level observed in a mock-transfected sample and with the level of the WT sampled after treatment with the rimonabant inverse agonist. Thus, we refer to T3.46A as '*fully inactive*.' For the WT receptor in the absence of ligand the level of GTPγS binding is 91.3 fmol/mg, an increase of 37.8 fmol/mg above the level of the mock-transfected sample whereas for the T3.46I mutant it is 112.4 fmol/mg, an

increase of 58.9 fmol/mg. These data are consistent with the previous ligand binding profile (2), confirming that the T3.46A receptor is inactive and unable to bind to G protein, the WT receptor has constitutive activity, and the T3.46I receptor is much more constitutively active, with the latter two able to couple to the G protein.

To provide the structural information needed to understand activation of GPCRs, we used first principles-based computational methods (4-7) to predict the ensemble of energetically favorable conformations available to play a role in multiple GPCR-mediated pathways. Our objective is to accurately predict multiple conformations of a GPCR that may or may not be closely related to the crystallographically characterized receptors. Homology modeling approaches are not yet capable of predicting such multiple conformations, as shown below in the discussion. We report here the results of applying these methods to CB1, providing a structural basis for understanding the dramatic effects on activation observed experimentally for single mutations at the T3.46 position in this receptor (2). We then use this structural understanding to predict two double mutants expected to convert the fully inactive T3.46A to WT-level constitutive activity, which we subsequently confirm experimentally by GTPγS binding assays. Conversely, we design another double mutant to convert the highly constitutively active L3.43A mutant to T3.43A-level full inactivity. To prove that a coupling involving the R2.37 residue maintains the inactive structure, we designed a triple mutant predicted to be constitutively active. Subsequent GTPγS assays agree with the functional levels and strongly support the predicted structures, thus reaffirming the proposed structural basis for the activation mechanism for the CB1 receptor.

METHODS

(For more details about structure prediction methodology, see *Chapter II*.)

To identify a structural basis for the experimental findings, we used the GEnSeMBLE method (4-8) to predict the ensemble of energetically favorable conformations, in other words the packing within the seven helix bundle, for all CB1 receptors discussed herein.

Part A: Predicting the WT receptor structures

Step 1: Predict transmembrane (TM) helix regions of the receptor

Using a BLAST (9) search, we found 1387 sequences of seven-helix proteins with sequence identities ranging in from 100% to 8.47%. **Table 3.1A** compares the sequence identities of the crystallized GPCR sequences to the human CB1 receptor sequence. The overall sequence identities of the seven crystal structures fall in the ~10-16% range with TM sequence identities ranging between ~16-28%. The other member of the human cannabinoid family, CB2, has a 32.42% overall sequence identity with human CB1 and a 52.59% sequence identity within the CB1 TM regions. The human CB1 receptor lacks certain motifs common to other class A GPCRs such as the conserved proline in TM5 and a disulfide bond between TM3 and the second extracellular loop (EC2). Thus, homology modeling is not expected to be effective as demonstrated by the total energy comparison discussed below in Part A Step 4. Moreover, our results show dramatic changes in TM orientations with just a single point mutation, invalidating the fundamental assumption in homology modeling.

Table 3.1 Constructing the WT, T3.46A, T3.46I receptors using first principles-based methods.
(A) Comparison to the human CB1 receptor sequence of the overall sequence identity and TM sequence identity of human CB2 and crystallized GPCR sequences including the human adenosine A_{2A} receptor ($HA_{2A}AR$) (10), the turkey β_1 adrenergic receptor ($T\beta_1AR$) (11), the human β_2 adrenergic receptor ($H\beta_2AR$) (12), the squid rhodopsin receptor (sRho) (13; 14), the human dopamine D3 receptor (hDD3R) (15), the bovine rhodopsin receptor (bRho) (16), and the human chemokine CXCR4 receptor (hCXCR4) (17). **(B)** Sampling rotation angles for CB1 receptors. Optimized eta (η, rotation), angles from the ComBiHelix program for the three receptors: T3.46A, WT, and T3.46I. The crystal inact$T\beta_1AR$ (PDB ID: 2VT4) (11) is the reference case. **(C)** Energy comparison for CB1 receptors with OptHelix helices and homologized helices. Optimized eta (η, rotation), angles from the ComBiHelix analysis for the WT receptors with various helices: OptHelix helices aligned to crystal inact$T\beta_1AR$ (PDB ID: 2VT4) template, homology inact$T\beta_1AR$ (PDB ID: 2VT4) helices, homology inact$H\beta_2AR$ (PDB ID: 2RH1) (18) helices, and homology inact$HA_{2A}AR$ (PDB ID: 3EML) (10) helices. * = Different selection of eta residue.

A.

GPCR Name	Overall Sequence Identity (%)	TM Sequence Identity (%)
Human CB1 Receptor	100	100
Human CB2 Receptor	32.42	52.59
Turkey Beta 1 Adrenergic Receptor	15.89	28.45
Human A_{2A} Adenosine Receptor	16.10	26.31
Human Beta 2 Adrenergic Receptor	14.19	24.65
Human D3 Dopamine Receptor	10.81	24.09
Bovine Rhodopsin Receptor	10.59	21.94
Human CXCR4 Receptor	9.96	18.57
Squid Rhodopsin Receptor	13.14	16.27

B.

	η (°)						
Receptor	**H1**	**H2**	**H3**	**H4**	**H5**	**H6**	**H7**
T3.46A	-120	-150	0	0	-120	30	-90
WT	90	-60	0	0	-120	30	-120
T3.46I	150	-90	0	30	-120	0	-30

C.

	η (°)							PostMinE
Method	**H1**	**H2**	**H3**	**H4**	**H5**	**H6**	**H7**	**(kcal/mol)**
Gensemble— inact$T\beta_1AR$ template	90	-60	0	0	-120	30	-120	-110.4
Homologized inact$T\beta_1AR$ Helices	-30	0	0	-30	-210*	0	90	108.1
Homologized inact$H\beta_2AR$ Helices	0	0	0	0	-90*	0	90	15.9
Homologized inact$HA_{2A}AR$ Helices	0	0	0	-30	90*	90	-120	109.6

In the second step, we aligned the 1387 amino acid sequences and found an average hydrophobicity value for each residue. The CB1 hydrophobicity values are plotted in **Figure 3.2**, and we can clearly identify the seven hydrophobic regions above the 0 baseline corresponding to the seven TM helical domains. The midpoint of the hydrophobic region for each helix represents the point where the helix crosses the hydrophobic plane passing through the center of the helical bundle.

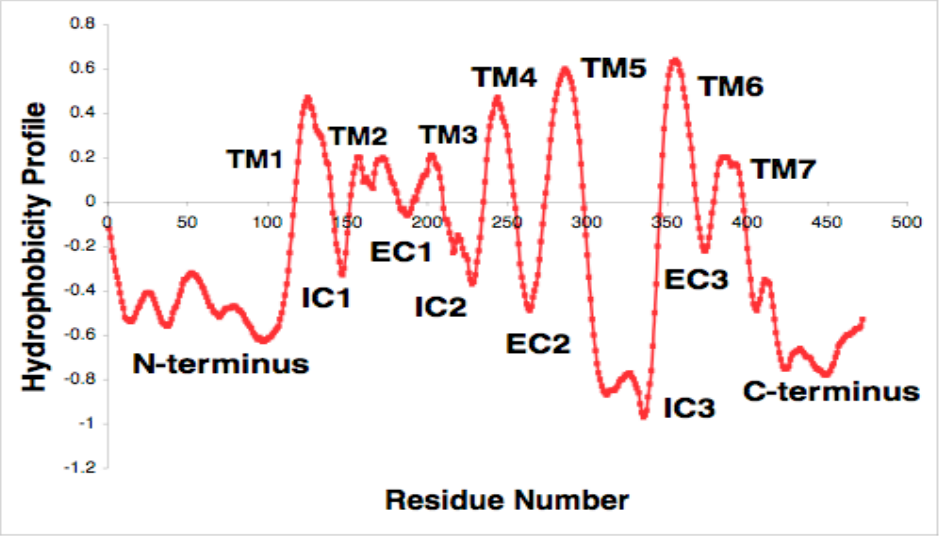


Figure 3.2 CB1 hydrophobicity plot. This graph indicates the hydrophobic portions of the amino acid sequence, which include the hydrophobicity values above the baseline. The seven hydrophobic regions correspond to the seven TM regions of the amino acid sequence.

Step 2: Optmize helix shapes for the seven TM segments using OptHelix

Figure 3.3 shows the OptHelix products for TMs 6 and 7. These helices contain proline residues, which impact the backbone shape by altering the bend, wobble, and face shift angles, as defined in Ref. (19) and indicated in the figure. The P6.50 residue (**Figure 3.3A**) causes the helix to bend 11.2°. The wobble angle, defined to be orientation of the post-proline portion of the helix with respect to the axis containing the proline's alpha-carbon, is a significant -128.7°. The face shift is the angle between proline's alpha-carbon and the vector between the alpha-carbons of the (*i*-3) and (*i*-4) residues away from proline indicating if the helix is over- or under-wound. The positive 48.1° face shift angle for P6.50 shows that the helix is under-wound. P7.50 (**Figure 3.3B**) causes a less significant bend of only 6.7°. TM7's wobble angle is at 167.5°, which is in the opposite direction of the one in TM6. The face shift angle is less significant in TM7 at 22.1°, but like TM6, this helix is under-wound.

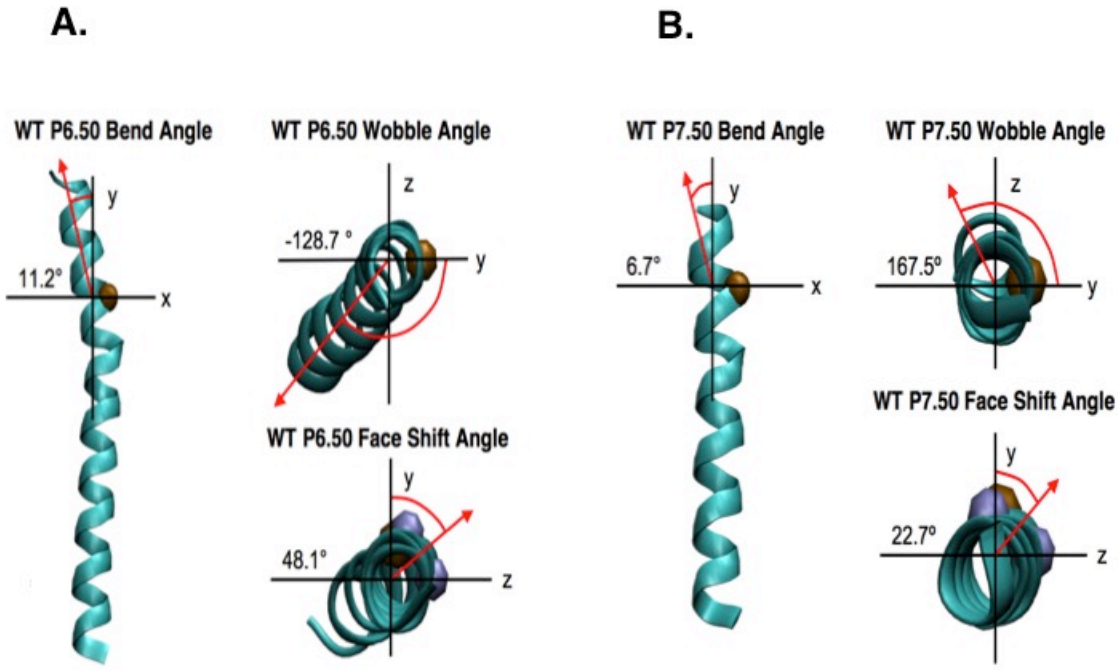


Figure 3.3 Effect of the proline residue on the shapes of **(A)** TM6 and **(B)** TM7. The helical backbone is shown in cyan, while the proline alpha-carbon atoms are brown spheres, and the (*i*-3) and (*i*-4) residue alpha-carbon's are lavender spheres. The axes and bend, wobble, and face shift angles are labeled in the same way as in Ref. (19).

Step 3: Generate template for receptor bundle

The seven optimized helices are then aligned to the experimental crystal templates inact$T\beta_1AR$ (PDB ID: 2VT4), (PDB ID: 2RH1), and inact$HA_{2A}AR$ (PDB ID: 3EML). The inact$T\beta_1AR$ template was chosen over the other two templates because our calculations gave lower energies for the inact$T\beta_1AR$ bundles after the ComBiHelix stage.

Step 4: Create ensemble of orientations of seven helices

Out of 1000 bundles generated by ComBiHelix (6), we select the lowest energy bundle, as shown in the **Table 3.1B**, which will be used by SuperBiHelix (8) as the starting point for sampling the tilt angles (θ and φ) simultaneously with the η angles. After which, we select the top 2000 based on lowest total SuperBiHelix combinations of pair-wise energies. These are built into full seven helix bundles in SuperComBiHelix. Then we evaluate the full energy of each of these 2000 conformations after optimizing side chains and minimizing the energy.

In addition to the above predictions, we repeated this procedure using helices from crystal homology models including inact$T\beta_1AR$ (PDB ID: 2VT4) (11), inact$H\beta_2AR$ (PDB ID: 2RH1) (18), and inact$HA_{2A}AR$ (PDB ID: 3EML) (10). **Table 3.1C** shows the final energy comparison

of the WT receptor built from various helices. The OptHelix helices produced the lowest energy helix bundle with a post-minimization energy value of -110.4 kcal/mol. The three cases with homologized helices range in energies of 15.9 to 109.6 kcal/mol, which are ~100-200 kcal/mol worse than that of the one with OptHelix helices. Furthermore, we find the homologized helices to be physically unreasonable because in the homologized helix structures, rimonabant-binding residues W5.43 and/or K3.28 (11; 20) are in positions that are not accessible to the binding site.

Step 5: Choose the lowest energy structures

We chose an ensemble of 10 low average energy rank poses that represent the most favorable and energetically accessible conformations.

Part B: Predicting the T3.46A, T3.46I, T3.46A/R2.37A, T3.46A/R2.37Q, L3.43A, L3.43A/D2.63A, L3.43A/D2.63A/R2.37A, and L3.43A/K3.28A receptor structures

Mutate:

- T3.46 residue into alanine or isoleucine and/or R2.37 into alanine or glutamine
- L3.43 residue into alanine and/or D2.63 into alanine
- L3.43 residue into alanine, D2.63 into alanine, and R2.37 into alanine
- L3.43 residue into alanine and K3.28 into alanine

using the Side Chain Rotamer Excitation Analysis Method (SCREAM) (21) in the OptHelix helices then repeat the subsequent steps outlined above for the WT receptor. Sometimes multiple low energy structures from the ComBiHelix stage were sources of input into SuperBiHelix, and all the resulting SuperComBiHelix structures were combined and ranked by average energy rank.

RESULTS

Position of the mutated residue T3.46

The T3.46 residue (T210) is one helical turn above R3.50 (see **Figure 3.4A**), which is highly conserved among class A GPCRs. In the WT CB1 receptor, R3.50 makes a critical salt bridge to D6.30 (D338). The crystal structures of bRho (16; 22-31), sRho (13; 14), and hDD3R (15) and other experimental studies (32; 33) have concluded that in the inactive form R3.50 forms a salt bridge with D/E6.30 on the cytoplasmic end of TM6, leading to the proposal that this salt bridge stabilizes the inactive form of the receptors and hinders their activation.

Our computational predictions indicate that WT CB1 contains a unique hydrogen bond involving T3.46 and S2.45 (**Figure 3.4B**). The S2.45 residue usually participates in the highly conserved TM2-TM3-TM4 hydrogen bond network as observed in other crystallized GPCRs, but these receptors have alanine or leucine at this position, so for all other receptors, this hydrogen bond connecting TM2 and TM3 cannot exist. By mutating this particular residue to alanine or isoleucine the hydrogen bond is broken, affecting the interactions between the TM regions. The changes in the receptor conformation are reflected in the changes in experimental GTPγS binding shown in **Figure 3.1** and will be discussed further below.

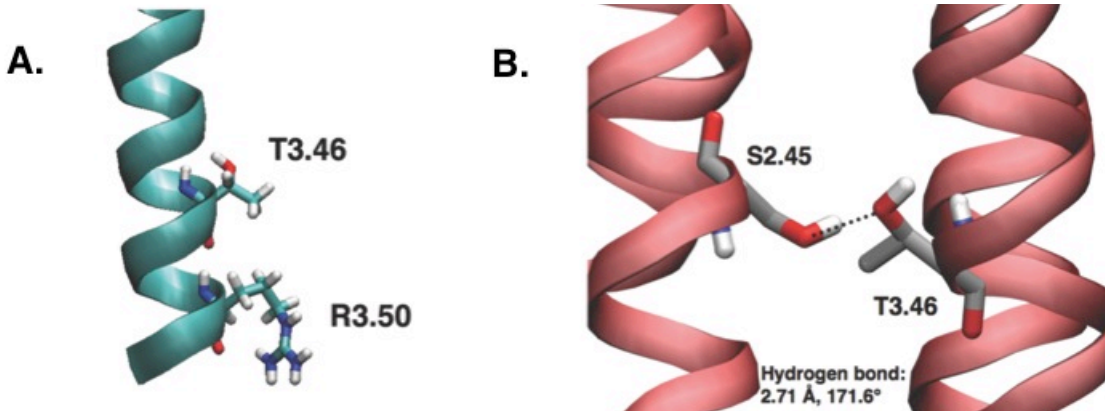


Figure 3.4 Relative position of T3.46 in the WT receptor. **(A)** Predicted structure of TM3 of human CB1 receptor's helix 3. Residues R3.50 and T3.46 (indicated) are on the same face of the helix, pointing into the center of the receptor bundle. **(B)** Atomic view of the interaction between T3.46 and S2.45 in WT CB1. Interaction between T3.46 and S2.45 in WT-1 (lowest average energy rank) conformation. The two residues interact via a hydrogen bond. The heteroatom distance is indicated by the dotted line. The S2.45-T3.46 hydrogen bond was found in five of our ten lowest energy protein structures after molecular dynamics (MD) annealing (34) on just the side chains of the two residues. This hydrogen bond is especially important because it involves the residue, T3.46, selected for mutation.

Predictions of the stable structural ensembles for CB1 WT and mutant receptors

We find that the mutations of T3.46 lead to substantial changes in helix packing within the seven-helix bundle, which in turn lead to major differences in the salt bridge interactions between the cytoplasmic end of TM6 that are expected to play a role in activation. The predicted ensemble of low energy structures of WT CB1 and its two mutants, T3.46A and T3.46I, are shown in **Table 3.2A**. The orientations of some helices change dramatically upon mutation. Here, we denote the angle about the helix axis as eta (η, rotation) and the two angles defining the tilt as theta (θ, tilt) and phi (φ, sweep) as defined in Ref. (35). The average energy ranges are:

- T3.46A: -334.3 kcal/mol to -311.2 kcal/mol,
- WT: -350.8 kcal/mol to -332.0 kcal/mol,

- T3.46I: -317.1 kcal/mol to -289.9 kcal/mol.

And the backbone root-mean-square deviation (RMSD), with respect to the lowest energy structure, ranges are:

- T3.46A: 0.2 Å - 3.5 Å,
- WT: 0.4 Å - 2.1 Å,
- T3.46I: 0.9 Å - 2.4 Å.

The lowest energy conformations for each receptor are structurally similar to one another as indicated by their RMSDs. For the crystallized GPCRs, their crystal resolution ranges from 1.80 Å (36) to 4.15 Å (28). All of the calculated RMSD values are either lower than or fall within the crystal resolution range indicating that x-ray crystallography could not distinguish the various low energy conformations for a particular receptor. These predicted structures do, however, differ greatly from the x-ray crystal structure of inact$T\beta_1AR$ (PDB ID: 2VT4) (11). Thus homology modeling would be ineffective for predicting CB1 receptor conformations, especially for the active states.

Table 3.2 Comparison of the conformations of the low-lying energy structures for the three receptors T3.46A, WT, and T3.46I. Theta (θ, tilt), phi (φ, sweep), eta (η, rotation) angles and the backbone RMSD (Å) with respect to the lowest energy conformation of the ten lowest energy T3.46A, WT, and T3.46I receptors are shown below. The inact$T\beta_1AR$ crystal (PDB ID: 2VT4) (11) is the reference case for the angles. The poses were chosen according to their average energy rank: the average of the charged total energy rank, neutralized total energy rank, charged interhelical energy rank, and neutralized interhelical energy rank.

Name	θ (°)							φ (°)							η (°)							Ave E	RMSD
	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	kcal/mol	Å
T3.46A-1	10	0	-10	10	-10	0	0	-30	0	-30	0	15	30	0	-120	-150	-15	30	-150	15	-90	-334.3	0.0
T3.46A-2	10	0	-10	10	-10	0	0	-30	0	-30	0	15	30	0	-105	-150	-15	30	-150	15	-90	-323.9	0.2
T3.46A-3	0	-10	-10	10	10	0	0	-30	-30	15	-30	-15	15	-15	-135	-165	-15	15	-105	45	-60	-328.0	2.6
T3.46A-4	0	-10	-10	10	10	0	0	-30	-15	30	-30	0	0	-15	-150	-180	0	-15	-105	45	-60	-318.2	3.5
T3.46A-5	0	0	-10	10	-10	0	0	0	0	-30	0	15	30	-30	-105	-150	-15	30	-150	15	-105	-323.5	1.8
T3.46A-6	0	-10	-10	10	10	0	0	-30	-30	30	-30	0	0	-15	-135	-150	-15	15	-105	45	-60	-315.9	3.5
T3.46A-7	0	0	-10	10	0	0	0	0	0	-15	0	15	15	-30	-105	-150	-15	30	-90	0	-105	-311.2	2.4
T3.46A-8	0	-10	-10	10	0	0	0	-30	-30	15	-30	-15	15	-30	-135	-165	-15	15	-105	45	-60	-316.6	2.7
T3.46A-9	10	-10	-10	10	0	0	0	-30	-30	15	-30	-15	15	-15	-135	-165	-15	15	-105	45	-60	-328.0	2.5
T3.46A-10	10	0	-10	10	-10	0	0	-30	0	-30	15	15	30	0	-120	-150	-15	30	-150	15	-90	-313.1	0.3
WT-1	0	-10	-10	10	0	-10	-10	-30	-15	15	-15	-15	-15	-15	90	-60	0	0	-150	15	-105	-344.8	0.0
WT-2	-10	-10	-10	10	0	0	0	-30	-30	15	-30	15	0	0	60	-75	0	15	-150	15	-105	-350.4	2.1
WT-3	-10	-10	-10	10	0	0	0	-30	0	15	-30	-15	0	-15	105	-60	0	0	-150	15	-105	-345.7	1.7
WT-4	0	-10	-10	10	0	0	0	-30	-15	15	-15	-15	0	-30	75	-60	0	0	-150	15	-105	-335.5	1.5
WT-5	-10	-10	-10	10	0	-10	-10	-30	-15	15	-30	-15	30	0	60	-75	0	15	-150	15	-105	-337.6	1.3
WT-6	0	-10	-10	10	0	-10	-10	-30	-15	15	-15	-15	-30	-30	75	-60	0	0	-150	15	-105	-336.7	0.4
WT-7	0	-10	-10	10	0	0	0	-30	15	15	-30	-15	0	0	90	-30	0	15	-150	15	-105	-338.8	1.7
WT-8	-10	-10	-10	10	0	0	0	-30	-30	15	-30	-15	0	0	60	-75	0	15	-150	15	-105	-345.8	1.8
WT-9	0	-10	-10	10	0	0	0	-30	-15	15	-30	-15	0	-15	105	-75	0	0	-150	15	-105	-332.4	1.5
WT-10	-10	-10	-10	10	0	0	0	-30	-15	15	-30	-15	0	-15	60	-75	0	0	-150	15	-105	-332.0	1.7
T3.46I-1	-10	-10	-10	10	0	-10	-10	-30	30	0	-15	-15	-15	-30	180	-90	0	30	-150	30	-60	-304.1	0.0
T3.46I-2	-10	-10	-10	10	0	-10	-10	-30	0	0	-15	-15	15	-15	180	-60	0	30	-150	15	-45	-305.6	0.9
T3.46I-3	-10	-10	-10	10	0	-10	-10	-30	0	0	-30	0	-15	-30	180	-60	0	30	-90	-15	0	-298.7	1.6
T3.46I-4	-10	-10	-10	10	0	-10	-10	-30	0	0	-15	0	-15	-15	120	-60	0	30	-90	-15	0	-295.1	1.7
T3.46I-5	-10	-10	-10	10	10	-10	-10	-30	0	15	-30	-15	-30	0	180	-75	-30	30	-90	-30	-15	-317.1	2.1
T3.46I-6	-10	-10	-10	10	0	-10	-10	-30	0	0	-15	-15	15	-30	180	-60	0	30	-150	15	0	-297.6	1.1
T3.46I-7	-10	-10	-10	10	0	-10	-10	-15	-15	15	-30	-15	-15	-30	165	-90	0	30	-150	30	-60	-299.1	1.1
T3.46I-8	-10	-10	-10	10	0	-10	-10	-30	0	0	-15	0	-15	-15	180	-60	0	30	-90	-15	0	-294.2	1.6
T3.46I-9	-10	-10	-10	10	0	-10	-10	-30	0	0	-15	-15	30	-30	180	-60	0	30	-90	0	0	-297.4	1.5
T3.46I-10	10	-10	-10	10	0	-10	-10	-30	0	0	-15	0	0	15	180	-60	0	30	-90	-30	-15	-289.9	2.4

Intermolecular interactions in the inactive T3.46A receptor

The consensus salt bridge patterns in the predicted ensemble of low energy conformations for the T3.46A, WT, and T3.46I receptors are shown in **Table 3.3** and **Figure 3.5.** The consensus set is defined by the presence of the particular interhelical interaction in more than half of the 10 lowest energy conformations. The differences in the consensus set of interactions for each receptor clearly show that there are major differences in their respective helical packings. The conformations of the inactive mutant T3.46A receptor lead to a consensus salt bridge pattern (**Table 3.3** and **Figure 3.5A**) that includes the following salt bridges, indicated by a plus (+) throughout the chapter, and hydrogen bonds, indicated by a dash (-) throughout the chapter:

R2.37+D6.30, R3.50+D6.30, D3.49+K4.41, D6.58+K7.32, H1.59-R4.39, H2.41-R6.32, and H2.65-K3.28.

Table 3.3 Comparison of consensus salt bridge patterns for each of the previously discussed receptors (T3.46A, WT, and T3.46I) plus our proposed double mutants (T3.46A/R2.37Q and T3.46A/R2.37A).

Receptor	Level of activity	R2.37(Q/A) + D6.30 hydrogen bond	R3.50 + D6.30 ionic lock	R2.37(Q/A) + D3.49 hydrogen bond	R5.71 + D6.30 salt bridge	D2.63 + K3.28 salt bridge	D3.49 + K4.41 salt bridge
T3.46A	Fully inactive	Yes	Yes	No	No	No	Yes
T3.46A/ R2.37A	Partially constitutively active	No	No	No	No	Yes	Yes
T3.46A/ R2.37Q	Partially constitutively active	No	Yes	Yes	No	Yes	Yes/ No
WT	Partially constitutively active	No	Yes	Yes	No	Yes	Yes
T3.46I	Highly constitutively active	No	No	Yes	Yes	Yes	No

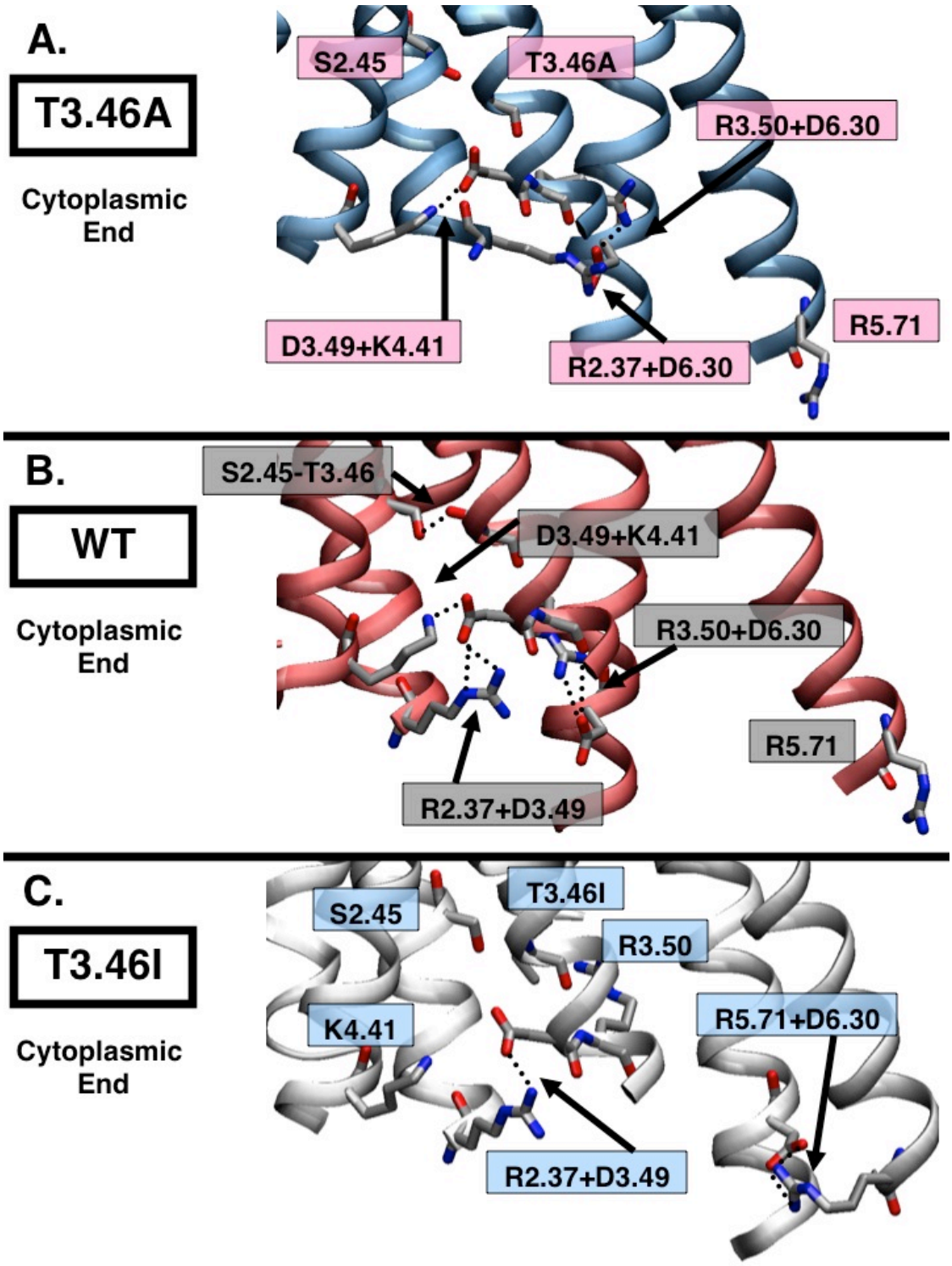


Figure 3.5A-C. Predicted structures of the T3.46A, WT, and T3.46I receptors, showing the salt bridges and hydrogen bonds formed on the cytoplasmic side. **(A-C)** Receptors are selected based on lowest average energy in **Table 3.2** and best agreement with the consensus pattern of salt bridges and hydrogen bonds. **(A)** T3.46A is represented by T3.46A-1, **(B)** WT is WT-3, and **(C)** T3.46I is T3.46I-4. These models for each receptor are referred to throughout this chapter. Salt bridges and hydrogen bonds in the consensus pattern are indicated by arrows and shaded boxes. Dotted lines show the hydrogen bonds between the residues.

We propose that the critical interactions opposing activation in this inactive T3.46A receptor are two ionic locks, defined as salt bridge interactions that putatively keep a receptor inactive:

- R3.50+D6.30 between the conserved residue arginine of the DRY motif at the cytoplasmic end of TM3 and the aspartic acid on TM6, and
- R2.37+D6.30 between arginine at the cytoplasmic end of TM2 and aspartic acid on TM6.

The hydrogen bond between TMs 2 and 6 (H2.41-R6.32, located approximately one turn above the R2.37+D6.30 interaction), further strengthens the coupling between TM2 and TM6.

The TM3+TM6 coupling is observed in crystal structures of such inactive GPCRs as bRho (16; 22-31), sRho (13; 14), hDD3R (15), and thermally stabilized inact$T\beta_1AR$ (37) and inact$HA_{2A}AR$ (38) mutants. Furthermore, experimental studies showed that protonation of E6.30 of the inact$H\beta_2AR$ causes activation (39), and that breaking of this ionic lock is necessary for full activation (40). The D3.49+K4.41 interaction is similar to the salt bridge between D3.49 and R4.37 in the hCXCR4 crystal (17) and the one possibly between D3.49 and R4.41 in hDD3R (15) (the side chain of R4.41 was not resolved in chain B of the crystal structure PDB ID: 3PBL). Likewise, the extracellular D6.58+K7.32 interaction is similar to the salt bridge between D6.58 and H7.32 in the hCXCR4 crystal (17).

On the other hand, the TM2+TM6 coupling is unique to this CB1 mutant. We believe this ionic lock is critical for maintaining the receptor in its fully inactive form.

Intermolecular interactions in the constitutively active WT receptor

The consensus interaction network for the WT conformational ensemble (**Table 3.3 and Figure 3.5B**) includes six salt bridges. Three of these salt bridges (D3.49+K4.41, R3.50+D6.30, and D6.58+K7.32) were also present in the T3.46A consensus and in the crystallographically characterized inactive class A GPCRs discussed above, and are therefore consistent with known systems.

The three additional salt bridges found in the WT protein conformational ensemble are R2.37+D3.49, H2.41+D3.49, and D2.63+K3.28. In the T3.46A mutant receptor, we found that R2.37 interacts with D6.30, but in the WT receptor, it forms a salt bridge with D3.49 as shown in **Figure 3.5B**. The most remarkable difference between the WT and T3.46A receptors is that the

T3.46A receptor has two polar interactions between TMs 2 and 6, R2.37+D6.30 and H2.41–R6.32, which are not observed in the WT receptor. Thus, instead of the direct TM2+TM6 coupling in the T3.46A mutant, the WT receptor displays the TM2+TM3 coupling between nearest-neighbor TMs.

The WT conformational ensemble consensus interaction also has four hydrogen bonds including S2.45-T3.46 (**Figure 3.4B**). This hydrogen bond is especially important because it involves the T3.46 residue, whose mutation causes the large changes in the structure being investigated here.

Intermolecular interactions in the more constitutively active (than WT) T3.46I receptor

The consensus of the highly constitutively active mutant T3.46I conformational ensemble (**Table 3.3** and **Figure 3.5C**) includes four salt bridges and two hydrogen bonds: R2.37+D3.49, H2.41+D3.49, D2.63+K3.28, R5.71+D6.30, Y2.40-R4.39, and C6.47-N7.45. The three salt bridges linking TMs 2 and 3 are present in the WT receptor, but none of them are in the T3.46A mutant. The R3.50+D6.30 interaction from the WT is broken in T3.46I as seen in **Table 3.3** and **Figure 3.5C**. Thus*, the T3.46I receptor has no salt bridges spanning the cytoplasmic side of the TM helix bundle*. This end of the T3.46I receptor has a larger diameter (11.6 Å between residues D6.30 and R3.50) than that of T3.46A (8.5 Å) or WT (8.9 Å) enabling it to potentially accommodate the C-terminal α-helix of the G_{α} protein during activation, which in act$H\beta_2AR$-G_s complex is 19.0 Å (PDB ID: 3SN6) (41). This makes the extracellular ligand binding pocket more accessible in the T3.46I mutant.

The R5.71+D6.30 interaction is unique to the T3.46I mutant conformation and is similar to the TM5+TM6 salt bridges observed in the constitutively active bRho mutant (PDB ID: 2X72) (31), the putatively active bOps (PDB ID: 3CAP, 3DQB) (42; 43), meta II (PDB ID: 3PQR, 3PXO) (44), and perhaps in a water-mediated interaction in the nanobody-bound activated act$H\beta_2AR$ (PDB ID: 3POG) (45) crystal structures. In active bRho, bOps, and meta II, the salt bridge is formed between K5.66 and E6.30. Thus, the T3.46I mutant has similar structural signatures observed in other activated GPCRs.

Effect of single amino acid mutations on CB1 receptor TM helix packing

Tables 3.4A-B highlight the differences in the helix orientation angles (defined in (6) and *Chapter II*) among the three receptors, also shown in **Figure 3.6**. The helical axis for these angles is uniquely defined for a straight or kinked helix by taking the least moment of inertia vector from

the eigensolution of the moment of inertia matrix for the helix obtained using only heavy backbone atoms (6). **Figure 3.6A** and **Table 3.4A** show that TM2 of WT rotates (η, helix rotation angle) toward TM3 by 90º while TM6 sweeps away (φ, helix sweep angle) from the center of the TM helix bundle clockwise by 30º. The combination of these two motions breaks the TM2+TM6 coupling of T3.46A. TM2 also tilts (θ, helix tilt angle) towards TM3 by 10º with respect to its counterpart in T3.46A. This movement plus its rotation allows the R2.37+D3.49 interaction to form. The TM3+TM6 coupling is maintained for both T3.46A and WT because TM3 sweeps towards TM6 by 45º. As for the D3.49+K4.41 interaction, TM4 of the WT rotates clockwise by 30º and sweeps 30º towards the center of the TM helix bundle compared to T3.46A. This forward-tilting motion of the cytoplasmic end of TM4 and the rotations of both the TM3 and TM4 helices help maintain the TM3+TM4 coupling in WT.

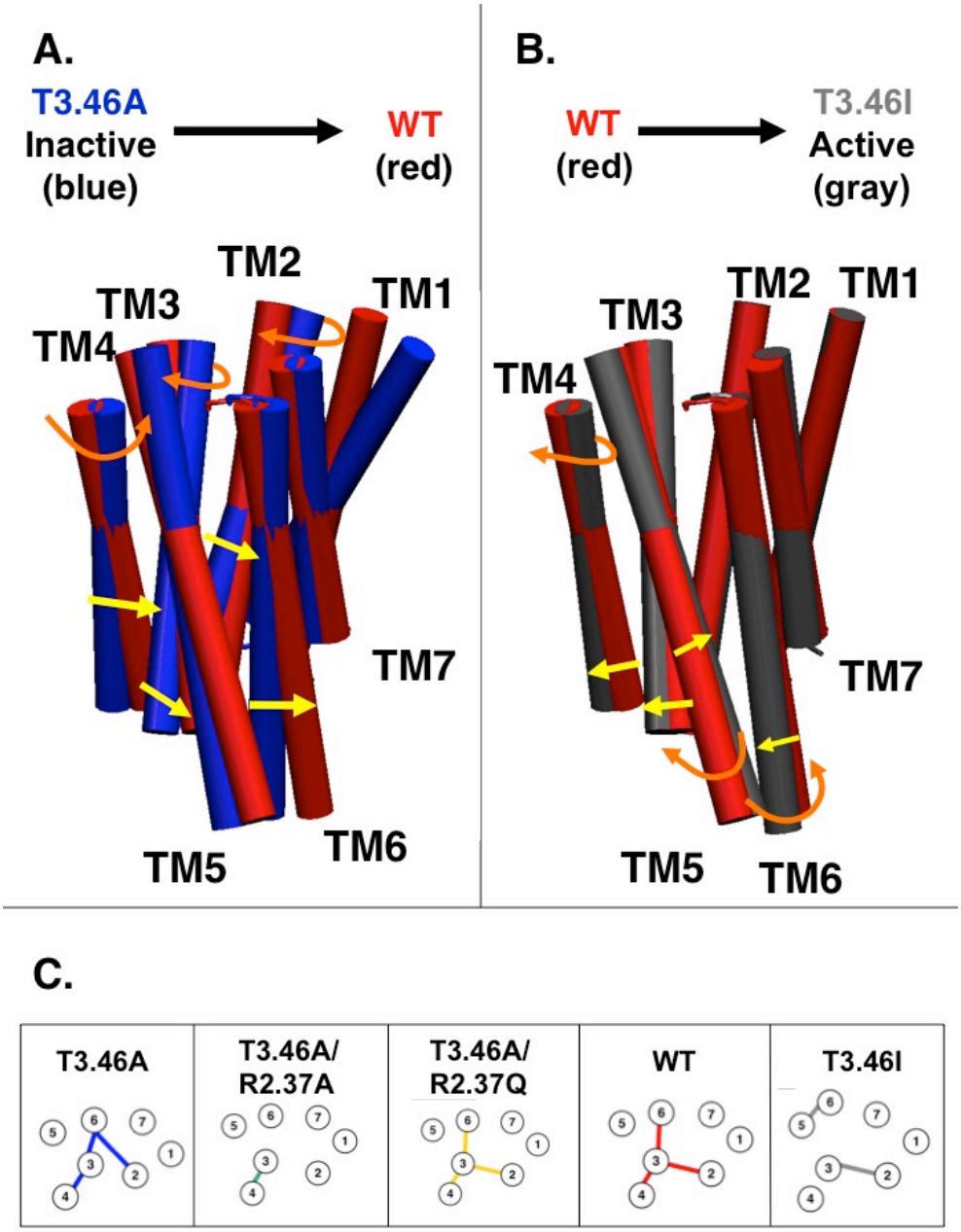


Figure 3.6A-C. Conformational changes in CB1 receptors during activation and the effect on salt bridges. **(A-B)** Comparisons of TM helix bundle rearrangement caused by the mutation of **(A)** T3.46A (blue) to WT (red), or **(B)** WT to T3.46I (grey). The structures chosen to illustrate the activation mechanism are based upon the lowest average energy rank and the salt bridge patterns shown in **Figures 3.5A-C**. The T3.46A model corresponds to T3.46A-1, the WT model matches WT-3, and the T3.46I's counterpart is T3.46I-4. Arrows indicate the direction of movement and the change is specified as shifts in the tilt (θ) [yellow], sweep (φ) [yellow], or rotation (η) [orange] angles. **(C)** The ionic locks and salt bridges critical for activation are illustrated in the predicted structures of human CB1. Circles represent the TM regions (numbered accordingly). Lines connecting the TMs indicate the intracellular salt bridges stabilizing the alpha-helices. Extracellular salt bridges have been omitted. Lines represent the salt bridges present in the following receptors: T3.46A (blue), T3.46A/R2.37A (green), T3.46A/R2.37Q (yellow), WT (red), and T3.46I (grey).

Table 3.4A-B. The predicted structures for the T3.46A and T3.46I mutants have different helix packing angles. **(A-B)** Differences in θ, φ, and η angles for the **(A)** T3.46A and WT structures and **(B)** WT and T3.46I structures, which represent the two steps of activation.

A.

	θ (°)				φ (°)				η (°)			
TM#	**2**	**3**	**4**	**6**	**2**	**3**	**4**	**6**	**2**	**3**	**4**	**6**
T3.46A-1	0	-10	10	0	0	-30	0	30	-150	-15	30	15
WT-3	-10	-10	10	0	0	15	-30	0	-60	0	0	15
Change (T3.46A-WT)	10	0	0	0	0	-45	30	30	-90	-15	30	0

B.

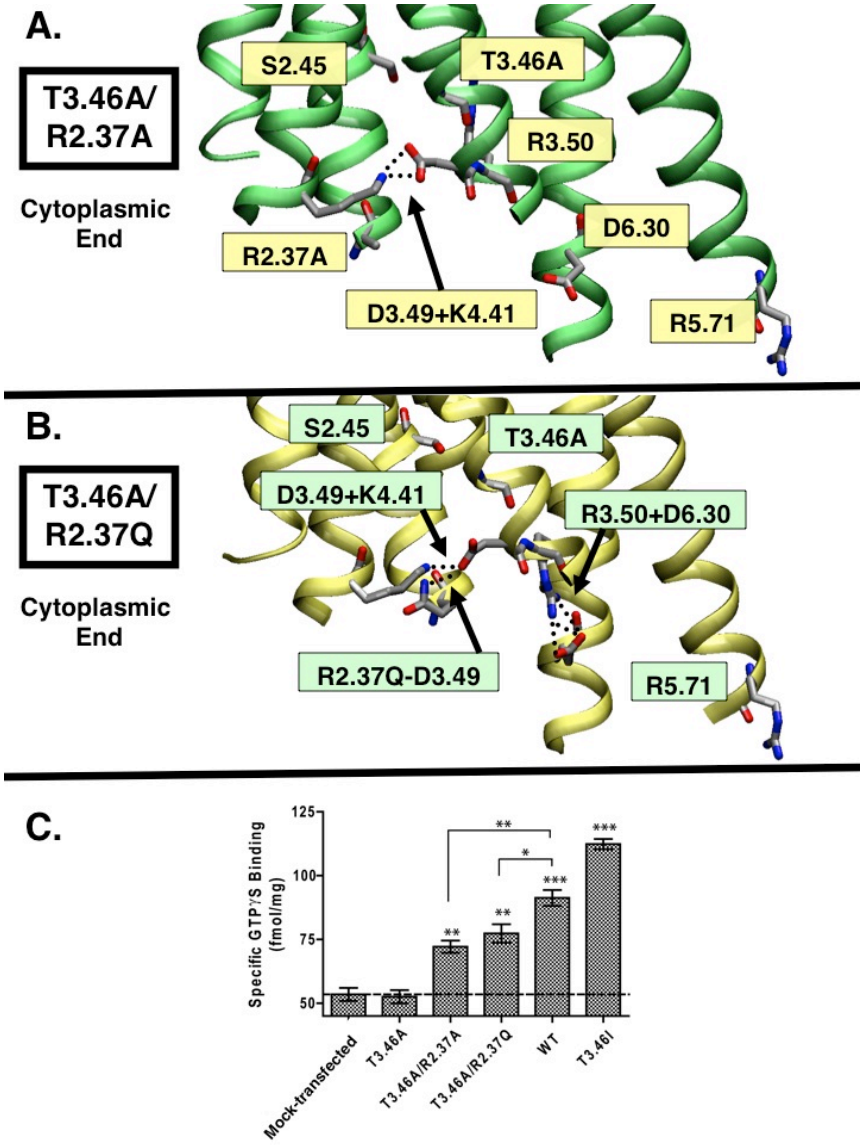
	θ (°)					φ (°)					η (°)				
TM#	**2**	**3**	**4**	**5**	**6**	**2**	**3**	**4**	**5**	**6**	**2**	**3**	**4**	**5**	**6**
WT-3	-10	-10	10	0	0	0	15	-30	-15	0	-60	0	0	-150	15
T3.46I-4	-10	-10	10	0	-10	0	0	-15	0	-15	-60	0	30	-90	0
Change (WT-T3.46I)	0	0	0	0	10	0	15	-15	-15	15	0	0	-30	-60	30

Comparing T3.46I with WT (**Table 3.4B** and **Figure 3.6B**), we see that TM6 tilts 10° away from the center of the bundle, breaking the TM3+TM6 coupling. In order to form the TM5+TM6 coupling of T3.46I, the extracellular ends of TMs 5 and 6 sweep away from each other by 15° causing their cytoplasmic ends to come closer together. These helices also rotate towards each other by 60° and 30°, respectively. The TM3+TM4 coupling is broken by the two helices leaning away from each other by 15°, while TM4 rotates by 30° away from TM3. The TM2+TM3 coupling is maintained due to the stationary TM2. This TM does not move despite the mutation of the adjacent TM's T3.46 to I3.46 whose bulkier side chain would clash with the residues on TM2. TM3 sweeps slightly away from TM2 by 15° to provide more space for the larger side chain of isoleucine and to avoid repulsive interactions.

Validation of structural basis of T3.46A inactivity by predicting a second mutation (at position R2.37) aimed at restoring constitutive activity to the T3.46A mutant

To evaluate the importance of the TM2+TM6 coupling in maintaining the inactive conformation of the receptor (as in T3.46A), we proposed two double mutants: T3.46A/R2.37A and T3.46A/R2.37Q **(Figures 3.7A** and **B)** that we predicted would restore constitutive activity by breaking the TM2+TM6 coupling. We expected that changing the arginine to a nonpolar alanine or to a charge-neutralizing glutamine on the cytoplasmic end of TM2 would break the R2.37+D6.30 interaction present in the inactive T3.46A receptor by weakening the interaction

with D6.30. This loss of TM2+TM6 coupling should cause the receptor conformation to change and resemble the constitutively active WT receptor, which should be observable in a functional assay.



Figures 3.7A-B. Conformational predictions of the constitutively active double mutants, T3.46A/R2.37Q and T3.46A/R2.37A. **(A-B)** Receptors are selected based on lowest average energy and best agreement with the consensus pattern of salt bridges and hydrogen bonds in **Table 3.3**. T3.46A/R2.37A **(A)** is represented by T3.46A/R2.37A-1, and T3.46A/R2.37Q **(B)** is T3.46A/R2.37Q-3. Salt bridges and hydrogen bonds in the consensus pattern are indicated by arrows and shaded boxes. Dotted lines show the hydrogen bonds between the residues. **(C)** Comparison of basal GTPγS binding to HEK293 cell membranes expressing the CB1 receptors including the double mutants (T3.46A/R2.37Q and T3.46A/R2.37A). The level of [^{35}S]GTPγS binding was measured in the absence of ligand for the WT, T3.46I, T3.46A, T3.46/R2.37A, and T3.46/R2.37Q receptors. The level of [^{35}S]GTPγS binding for the mock-transfected membrane sample is shown for comparison. The data description is same as for **Figure 3.1.**

We proceeded to predict the optimum ensemble of structures for these double mutants using the same computational procedure as for WT and the two single mutants. For T3.46A/R2.37A double mutant, the consensus interactions of the ten best-predicted conformations agree with our expectations, resembling a constitutively active receptor. All couplings between TM2 and TM6 and even the R3.50+D6.30 interaction are absent in the lowest energy conformations of this double mutant (**Figure 3.7A** and **Table 3.3**). It also lacks the R5.71+D6.30 interaction observed in the highly constitutively active T3.46I receptor. However, it does contain the D3.49+K4.41 interaction found in the T3.46A and WT receptors. This receptor also contains the D2.63+K3.28 interaction on the extracellular side of the receptor bundle. This salt bridge is observed in both the constitutively active WT and T3.46I receptors, but not in the inactive T3.46A receptor. The lack of the TM2+TM6 and TM3+TM6 couplings implies that this receptor would be constitutively active. The presence of the TM3+TM4 coupling as seen in WT and inactive T3.46A receptor suggests that this double mutant might be less constitutively active than the T3.46I receptor.

For T3.46A/R2.37Q double mutant receptor, the consensus of the ten best conformations agrees with our expectations, leading to a hydrogen bond and salt bridge pattern similar to that of the WT receptor (**Figure 3.7B** and **Table 3.3**). The glutamine residue at position 2.37 interacts with D3.49 as in the WT and T3.46I receptors, whereas in the inactive T3.46A receptor, the corresponding arginine interacts with D6.30. None of the low energy T3.46A/R2.37Q double mutant conformations show this interaction as expected. The R3.50+D6.30 interaction remains intact in this double mutant, appearing in eight of the ten lowest energy conformations. Like the T3.46A/R2.37A receptor, it too has the extracellular D2.63+K3.28 interaction, which as mentioned above, is seen only in the constitutively active receptors. Half of the structures have the intracellular D3.49+K4.41 interaction found in the T3.49A and WT receptors. Since the polar interaction patterns closely resemble those of the WT, we predict that the T3.46A/R2.37Q double mutant would make the inactive T3.46A mutant constitutively active. These predicted structures formed the basis of the subsequent GTPγS assays to test the activity of the double mutants as described below.

Experimental confirmation that the double mutants, T3.46A/R2.37Q and T3.46A/R2.37A, are constitutively active

Figure 3.7C shows a comparison of the level of basal GTPγS binding for the double mutants along with the other receptors and the mock-transfected sample shown in **Figure 3.1.** As

predicted, the T3.46A/R2.37A and T3.46A/R2.37Q receptors are more constitutively active relative to T3.46A. The GTPγS binding in the absence of ligand is larger for the T3.46A/R2.37A and T3.46A/R2.37Q receptors, with levels of 72.2 fmol/mg and 77.4 fmol/mg, respectively, than that of the fully inactive T3.46A receptor (52.6 fmol/mg). Although these two double mutants display less constitutive activity than the WT (91.3 fmol/mg), the mutations of the R2.37 residue produce substantial levels of constitutive activity. This confirms that our models are predictive and accurately show the importance of the critical TM2+TM6 ionic lock in keeping the T3.46A receptor in the inactive conformation. Intriguingly, these double mutant receptors show somewhat different non-covalent interaction patterns. The T3.46A/R2.37Q receptor has a R3.50+D6.30 interaction, whereas the T3.46A/R2.37A receptor does not (**Table 3.3**). However, our functional assay data indicate that the difference between the constitutive activities of these receptors is not statistically significant (**Figure 3.7C**).

CB1 has an ensemble of active states: L3.43A and T3.46I are both active, but have different receptor structures

Previous research suggests that the L3.43A mutant is highly constitutively active like the T3.46I mutant (46). In theory, the smaller alanine residue replaces the bulky leucine residue, which should allow enough space for the extracellular end of TM6 to tilt toward TM3. The L3.43A receptor has a slightly enhanced binding affinity for agonists and a significantly decreased binding affinity for the inverse agonists and antagonists when compared to the WT. L3.43A has a substantially enhanced GTPγS binding affinity of 114.5 fmol/mg, and T3.46I has a corresponding value of 112.3 fmol/mg, which correlates with increased activity.

We anticipated that the L3.43A and T3.46I would have similar conformations based on their similar levels of constitutive activity. However, **Figure 3.8** shows that there are major differences between the two receptors' structures. They have similar TM2 and TM3 contacts—both receptors have the extracellular D2.63+K3.28 and the intracellular R2.37+D3.49 salt bridges. Both of them lack the stabilizing TM2+TM6 and TM3+TM6 ionic locks as well as the TM3+TM4 salt bridge observed in the fully inactive or constitutively active mutants. However, the L3.43A mutant is missing the R5.71+D6.30 salt bridge that is characteristic in the highly constitutively active T3.46I mutant. TMs 5 and 6 in L3.43A are rotated so that it is impossible for the salt bridge contact to form. Our predicted structures indicate that the TM5+TM6 coupling is not necessary for activation to occur.

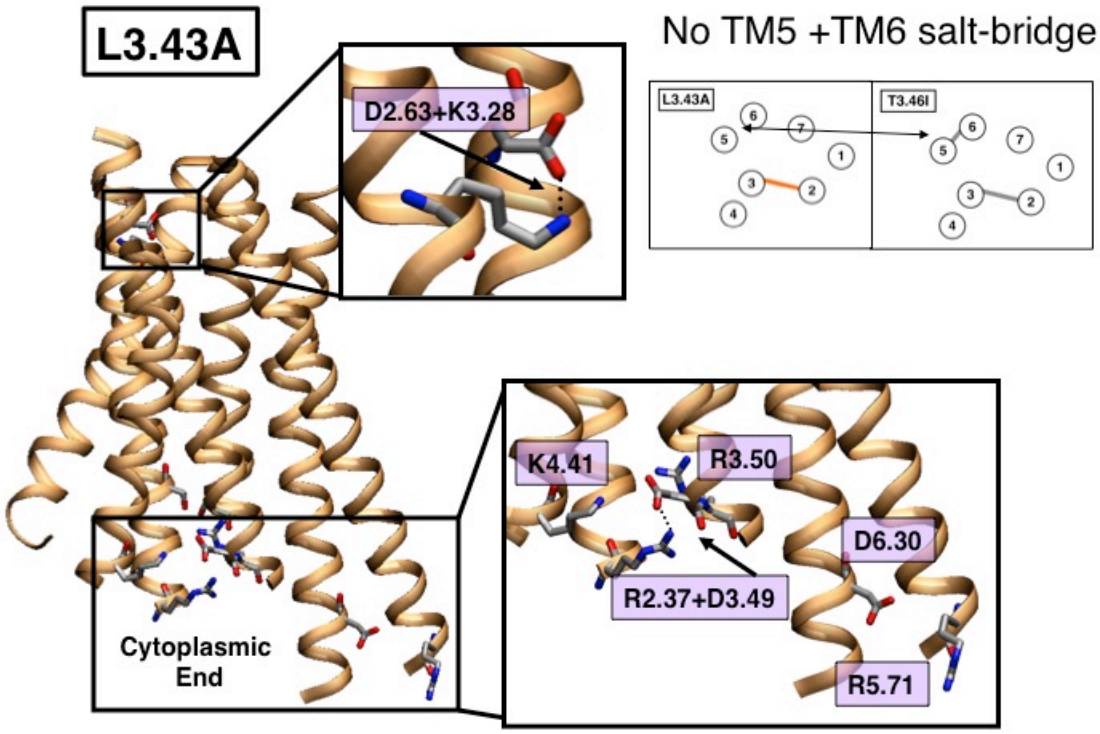


Figure 3.8 Predicted structure of the L3.43A receptor. Extracellular and intracellular salt bridges are indicated. L3.43A is missing the R5.71+D6.30 salt bridge found in the similarly highly constitutively active T3.46I mutant.

Validation of structural basis of L3.43A activity by predicting a second mutation (at position D2.63) aimed at eliminating constitutive activity to the L3.43A mutant

Previously, we showed that adding a second mutation to the fully inactive T3.46A causes the receptor to appear activated and experience a corresponding gain in function. In this section, we show that adding a second mutation causes a highly active mutant receptor to lose function. The L3.43A mutant has an extracellular salt bridge between D2.63 and K3.28, which is present in all of our receptors except the fully inactive T3.46A mutant, suggesting that this particular salt bridge is necessary for activation. In the L3.43A mutant, D2.63 was mutated into an alanine, and we predicted the structures of the resulting double mutant receptors.

The ensemble of ten low energy conformations of L3.43A/D2.63A shows a strong preference for the R2.37+D6.30 ionic lock formation that was observed in the fully inactive T3.46A mutant (**Figure 3.9**). Previously, we showed that this R2.37 residue is important for preventing activation, and the predicted structures suggest that it does so by forming a salt bridge with the D6.30 across the helix bundle. Surprisingly, only two out of the top ten L3.43A/D2.63A mutants had both TM2+TM6 and TM3+TM6 ionic locks, which differs from the salt bridge pattern

observed in the T3.46A receptor. The overall conformation of two fully inactive receptors is significantly different as their backbone RMSD is 2.32 Å, which is comparable to the difference between the two highly constitutively active T3.46I and L3.43A mutants. Furthermore, it calls into question the importance of the TM3+TM6 ionic lock for the fully inactive form of this receptor. These predicted structures for the two mutants suggest that there are multiple possible inactive conformations.

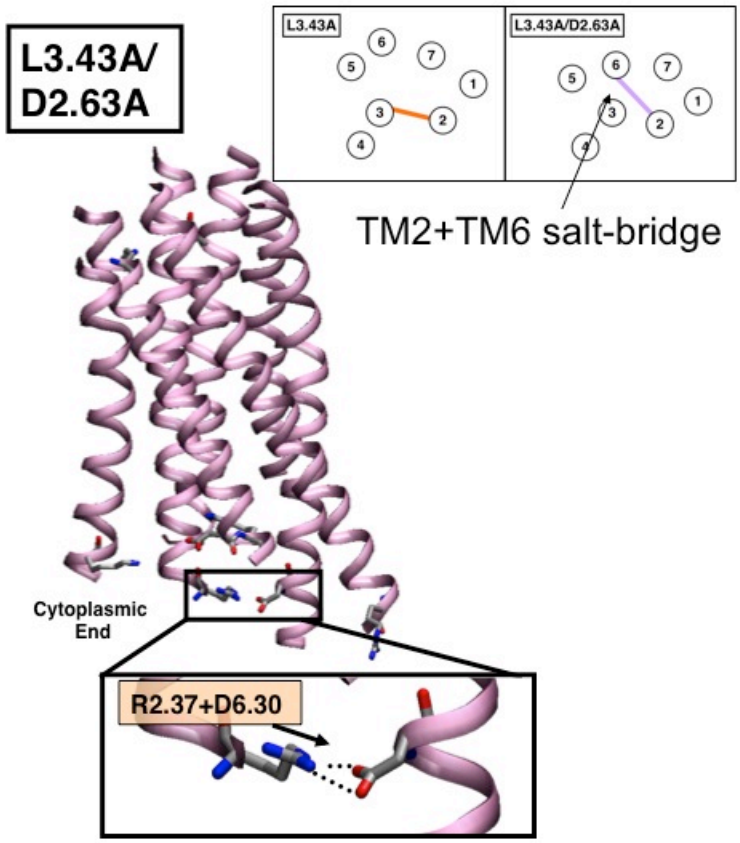


Figure 3.9 Predicted structure of the L3.43A/D2.63A receptor. The R2.37+D3.49 salt bridge observed in the L3.43A mutant was broken to form the new intracellular R2.37+D6.30 salt bridge (indicated).

Experimental confirmation that the double mutant, L3.43A/D2.63A, is fully inactive

Again, to confirm our structure and activity level predictions, the level of GTPγS binding was subsequently measured with assays to understand the relative levels of G protein binding and thus, the constitutive activity level (**Figure 3.10**). The L3.43A mutant had very high levels of GTPγS binding with a value of 114.5 fmol/mg, and the L3.43A/D2.63A mutant had significantly smaller levels of GTPγS binding with a value of 53.3 fmol/mg, which is approximately equal to the values for T3.46A (53.2 fmol/mg) and even the mock-transfected membrane (53.3 fmol/mg).

Thus, the L3.43A/D2.63A double mutant, much like the T3.46A one, is fully inactive possibly due to the presence of the TM2+TM6 coupling.

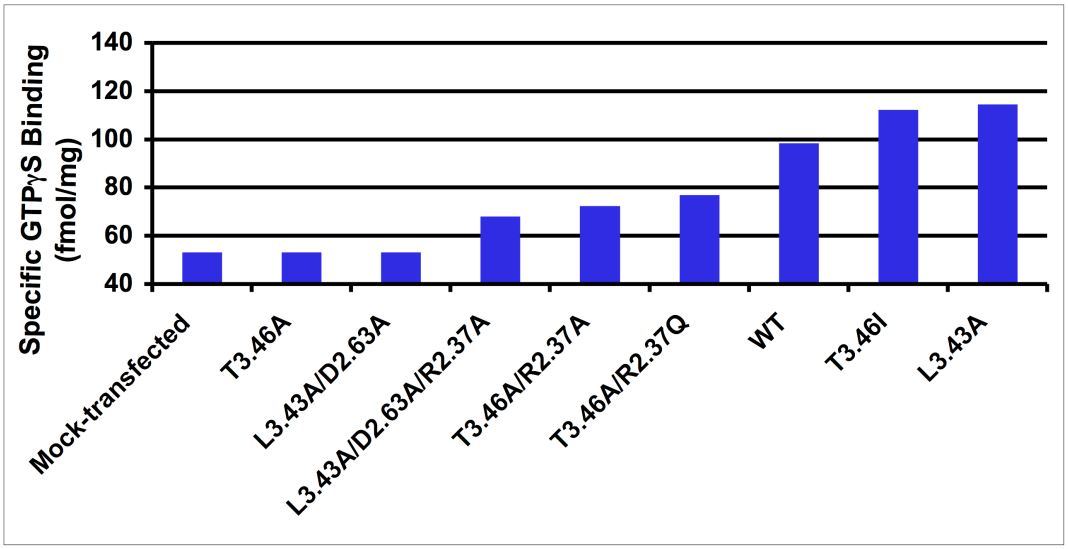


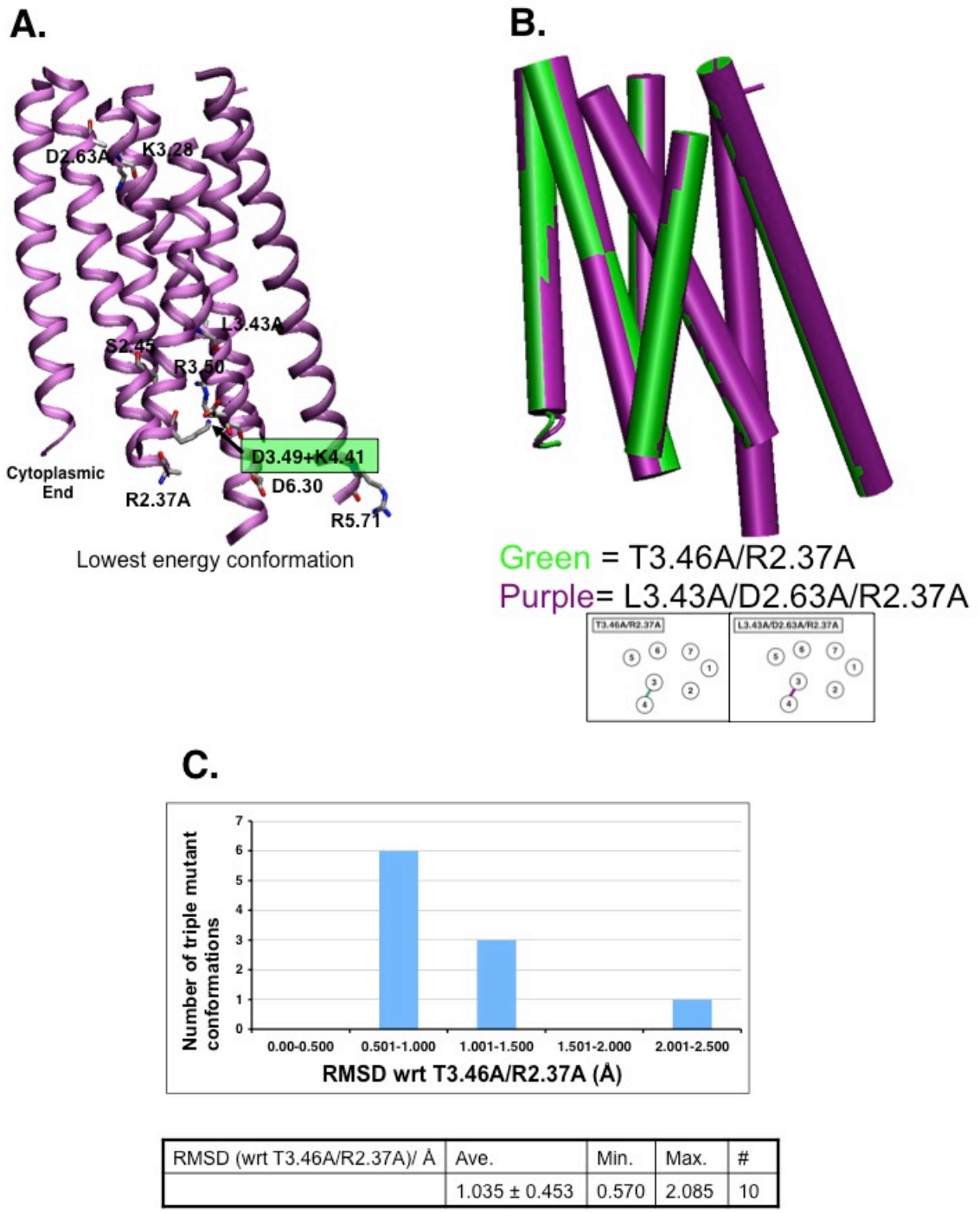
Figure 3.10 Comparison of basal GTPγS binding to HEK293 cell membranes expressing the CB1 receptors including the double mutant L3.43A/D2.63A and the triple mutant L3.43A/D2.63A/R2.37A). The level of [^{35}S]GTPγS binding was measured in the absence of ligand. The level of [^{35}S]GTPγS binding for the mock-transfected membrane sample is shown for comparison. The data description is same as for **Figure 3.1.**

Validation of structural basis of L3.43A/D2.63A inactivity by predicting a third mutation (at position R2.37) aimed at restoring constitutive activity to the L3.43A/D2.63A double mutant

Our predicted structure of the fully inactive double inactive mutant, L3.43A/D2.63A, has the characteristic TM2+TM6 coupling observed in the fully inactive T3.46A mutant. We have shown above that the R2.37 residue is necessary to stabilize the inactive conformation. We predict that this residue plays the same role in stabilizing the L3.43A/D2.63A double mutant. To confirm this hypothesis, we built the L3.43A/D2.63A/R2.37A receptor, a triple mutant. This is the inactive L3.43A/D2.63A double mutant with an extra mutation to break the coupling between TM2 and TM6.

Our predicted structures of the triple mutant look constitutively active based on observed salt bridge patterns (**Figure 3.11A**). The TM2+TM6 coupling was broken, and the R3.50+D6.30 ionic lock was also absent. Of the ten conformations in the low energy pool, seven of them showed the D3.49+K4.41 salt bridge, which was observed in the inactive T3.46A and constitutively active WT, T3.46A/R2.37A, and T3.46A/R2.37Q receptors, but neither of the highly constitutively active T3.46I and L3.43A receptors. The triple mutant also has the extracellular D6.58+K7.32 salt bridge, which was observed in the same receptors as the

TM3+TM4 salt bridge except for the constitutively active T3.46A/R2.37Q mutant. The consensus of the triple mutant shows that the majority of conformations lack the R5.71+D6.30 salt bridge observed in the highly constitutively active T3.46I mutant.



RMSD (wrt T3.46A/R2.37A)/ Å	Ave.	Min.	Max.	#
	1.035 ± 0.453	0.570	2.085	10

Figure 3.11 (A) Predicted structure of the triple mutant, the L3.43A/D2.63A/R2.37A receptor. The D3.49+K4.41 salt bridge is indicated. **(B)** Alignment of the lowest average energy conformations of the T3.46A/R2.37A (green) and L3.43A/D2.63A/R2.37A (purple) mutants. (**C**) Helical backbone RMSD analysis for the top ten lowest average energy conformations of L3.43A/D2.63A/R2.37A with respect to the lowest energy conformation of T3.46A/R2.37A. The histogram shows the number of triple mutant conformations within a given RMSD. The table shows the range and average of the ten RMSD values.

In fact, the triple mutant looks extremely similar to the T3.46A/R2.37A double mutant. They share the two salt bridges (D3.49+K4.41, D6.58+K7.32) that stabilize the receptor. **Figure 3.11B** shows the alignment of the two mutants and visually speaking, they are very similar and are almost identical in six of the seven TMs. **Figure 3.11C** shows how the top ten triple mutants ranked according to lowest average energy compare with the lowest energy T3.46A/R2.37A. The average backbone RMSD for the ten triple mutants is 1.04 Å, which is smaller than the lowest crystal resolution for a GPCR, 1.8 Å. The range of RMSD values is from 0.57 Å to 2.08 Å with nine of the ten structures having an RMSD of less than 1.50 Å. Based on the conformational similarities between the two receptors, we expect the triple mutant will be constitutively active like the WT and T3.46A/R2.37A receptors. Our structure predictions are confirmed with the GTPγS assays because the GTPγS binding level is 68.1 fmol/mg, which indicates that this triple mutant has a substantial level of constitutive activity (**Figure 3.10**). Its level of GTPγS binding is significantly less than that of the WT receptor (98.5 fmol/mg), but it is more than that of the fully inactive mutants (~53 fmol/mg) and approximately equal to that of the constitutively active double mutants T3.46A/R2.37A and T3.46A/R2.37Q, 72.5 fmol/mg and 77.0 fmol/mg respectively. This evidence shows that the R2.37 residue is important for preventing the activation of the inactive conformation of the L3.43A/D2.63A receptor.

Validation of structural basis of L3.43A activity by predicting a second mutation (at position K3.28) aimed at eliminating constitutive activity to the L3.43A mutant

Another way to test the importance of the D2.63+K3.28 salt bridge in stabilizing the active conformation is to mutate the K3.28 residue to break the TM2+TM3 coupling. Previously, when we mutated D2.63 to an alanine, the resulting conformation formed an R2.37+D6.30 ionic lock stabilizing the fully inactive conformation. If its partner is the K3.28 residue, than the L3.43A/K3.28A mutant should be fully inactive like the L3.43A/D2.63A mutant. The predicted structure of the L3.43A/K3.28A mutant (**Figure 3.12**) has multiple stabilizing salt bridges including the R2.37+D6.30 ionic lock, the R3.50+D6.30 ionic lock, and the extracellular D6.58+K7.32 salt bridge. The TM2+TM6 coupling was observed in the fully inactive T3.46A and L3.43A/D2.63A mutants, and it was experimentally proven that the R2.37 residue plays an important role in stabilizing the inactive state. Both the D6.58+K7.32 and R3.50+D6.30 salt bridges were observed in T3.46A. One discrepancy is that the majority of the L3.43A/K3.28A mutant conformations do not have the D3.49+K4.41 salt bridge, which was present in the inactive T3.46A and constitutively active T3.46A/R2.37A, T3.46A/R2.37Q, WT, and L3.43A/D2.63A/R2.37A receptors, but not the fully inactive L3.43A/D2.63A mutant. The

L3.43A/K3.28A mutant shown in **Figure 3.12** has the smallest average backbone RMSD, 2.14 Å, with respect to the L3.43A/D2.63A mutant, which is within the typical GPCR crystal resolution. Based on our structural predictions, we believe that L3.43A/K3.28 will be inactive like the L3.43A/D2.63A receptor as we had anticipated. However, we do not have experimental confirmation of our hypothesis. Site-directed mutagenesis, GTPγS assays, and ligand binding assays were not performed because previous experiments indicate K3.28 binds to ligands (20; 47; 48), so our results from experiments involving ligands would be complicated because we would not be able to tell if the physiological consequences were due to effects with the ligand or structural effects from to the broken D2.63+K3.28 salt bridge.

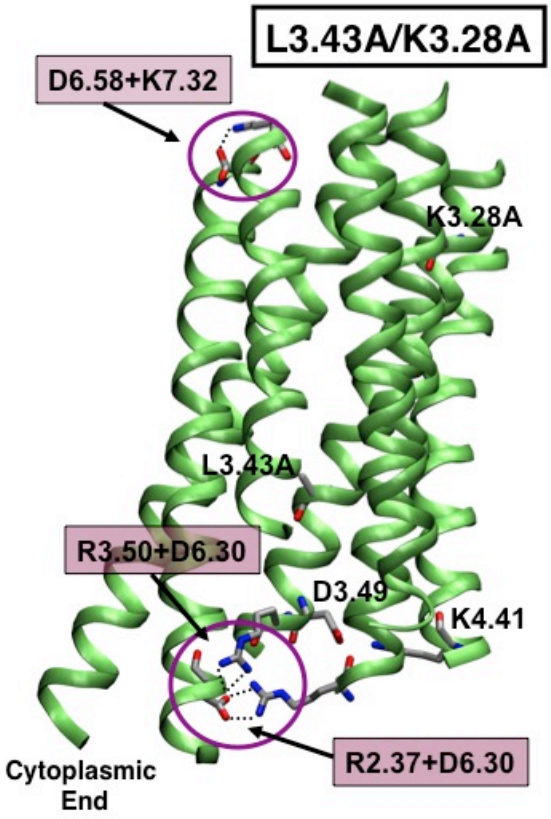


Figure 3.12 Predicted structure of the double mutant L3.43A/K3.28A. This conformation is the third lowest energy pose, which best represents the consensus of the ten lowest energy structures. The salt bridges are indicated with dotted black lines and arrows and are encircled in purple.

DISCUSSION

An activation mechanism for CB1 receptor consistent with crystallographically characterized GPCRs and mutagenesis experiments

By comparing the predicted salt bridge patterns for the human CB1 WT and two mutant receptors, T3.46A and T3.46I, we propose an activation mechanism in terms of receptor transformations. **Figure 3.5C** illustrates the predicted patterns of four salt bridges: R2.37+D6.30, R3.50+D6.30, D3.49+K4.41, and R5.71+D6.30.

- T3.46A (inactive) has the following salt bridges: R2.37+D6.30, D3.49+K4.41, and R3.50+D6.30.
- During the first step of partial activation to form the weakly constitutively active WT, the TM2+TM6 coupling in the inactive T3.46A receptor is broken to form the TM2+TM3 coupling of the WT receptor. Thus, the TM2+TM6 coupling appears to be essential for the receptor to be fully inactive.
- The TM3+TM4 and TM3+TM6 couplings break during the second step of activation to form the highly constitutively active T3.46I conformation with a new TM5+TM6 coupling. Thus, the breaking of the TM3+TM6 coupling appears to be responsible for full activation beyond WT levels.
- Our proposed double mutants, T3.46A/R2.37A and T3.46A/R2.37Q, share salt bridge signatures with the WT receptor. Based on their structures alone, we predicted that the receptors would be constitutively active, which was subsequently confirmed by experiments.

The CB1 receptor is unusual in that the TM3+TM6 coupling, R3.50+D6.30, appears to be insufficient for preventing activation. The lack of the TM3+TM6 coupling in the fully inactive double mutant L3.43A/D2.63A indicates that the TM2+TM6 ionic lock, not the traditional TM3+TM6 coupling, is necessary to stabilize the inactive CB1 conformation. Several studies have questioned the importance of that salt bridge interaction in the activation mechanism of the CB receptors. Site-directed mutagenesis of R3.50A in the closely related CB2 suggested that this mutation only weakly affects agonist binding and partially affects signaling (49). Another experiment showed that this same CB2 receptor mutation prevented constitutive activation and downfield signaling (50), which would contradict the previous study by indicating the importance of R3.50 in function and signaling. Both studies, however, do agree that the R3.50A mutation does not impact agonist binding. Our modeling studies provide a structural explanation of these experimental results. The R2.37+D6.30 ionic lock of the T3.46A mutant spanning the helix

bundle is unique while the R3.50+D6.30 ionic lock is commonly found in inactive GPCRs. R2.37 is present in CB1, but CB2 has a similar K2.37 residue that we suggest could form a similar TM2+TM6 coupling. In other class A GPCRs residue 2.37 is most often an asparagine, a serine, or a threonine, which explains why this particular ionic lock has not been observed in most crystalized receptors. The exception is the hS1P1R crystal structure, but its R2.37 side chain was not resolved in the crystal structure, so it is not known how this residue interacts. Indeed, it points into the center of the TM helix bundle making the TM2+TM6 coupling possible (51). This unique R2.37+D6.30 interaction rather than the R3.50+D6.30 ionic lock is likely responsible for preventing cannabinoid receptor activation, which would explain the limited importance of the R3.50 residue.

Validation of the role of TM2+TM6 interaction in keeping T3.46A inactive

The predicted structures of the CB1 receptor mutants are consistent with the experimental studies indicating that the R2.37+D6.30 interaction may be mandatory for the maintenance of the inactive state in the CB1 as well as CB2 receptors. To evaluate the importance of the TM2+TM6 coupling in maintaining the inactive conformation, we proposed two double mutants, T3.46A/R2.37A and T3.46A/R2.37Q, expected to break this coupling. We tested this prediction computationally and found that the TM2+TM6 coupling does not form. Our collaborators experimentally assessed the consequence of breaking this predicted R2.37+D6.30 interaction by evaluating the level of GTPγS binding, finding a substantial increase in constitutive activity relative to the T3.46A receptor. These results show that the double mutants are significantly more active than T3.46A just as we predicted.

These results underscore the importance of the R2.37 residue in maintaining the inactive conformation **(Figure 3.7C)**, confirming that the TM2+TM6 coupling is critical for constraining the receptor to maintain the inactive conformation. These experimental results further support our hypothesis that breaking the TM2+TM6, not the TM3+TM6, coupling is crucial for converting the inactive structure to the active conformation.

Comparisons to crystallized class A GPCRs

These conformational changes lead to helix movements consistent with those observed in the inactive-to-active experimentally characterized GPCRs. For example, they agree with NMR studies of bRho and meta II (52), which showed that TM3 remains fairly rigid and does not re-orient while TM6 moves significantly during activation. Recent crystal structures of activated

GPCRs allow us to compare the conformational differences between the antagonist and agonist-bound Hβ_2AR (41; 45) and $HA_{2A}AR$ (53), or the 11-*cis*-retinal-bound bRho and the apo-receptor bOps (43), with our weakly constitutively active WT and the highly constitutively active T3.46I, as seen in **Table 3.5.** TMs 1 - 4 are relatively fixed between bRho and bOps and inact$HA_{2A}AR$ and act$HA_{2A}AR$, whereas the TMs 5 - 7 show the greatest structural changes. For example, in bRho, TM6's cytoplasmic end moves 6 – 7 Å away from the bundle interior (43), and in $HA_{2A}AR$, the same helix shifts by 3 – 4 Å. $HA_{2A}AR$'s movement is smaller than that of bOps because $HA_{2A}AR$ may be constrained by the substitution of T4 lysozyme for the third intracellular loop (53).

Our CB1 structures show a similar change as TM6 swings outward by 6.6 Å, which agrees well with the bRho-bOps case (43). In Hβ_2AR, E6.30 moves by 11.40 Å when stabilized by a nanobody (45) or by 14.7 Å when stabilized by a G_s protein (41), while in CB1 the conserved residue shifts by 7.4 Å. Also evident in crystal structures is the movement of TMs 5 and 6 towards each other. In the bRho-bOps activation, the two helices tilt so they become parallel (43). We find the same shifts in TM5 and TM6 of CB1, as seen in **Figure 3.6B**. The most marked differences between our predicted CB1 structures and the crystallized GPCRs occur in the movement of TM7. None of the salt bridges crucial to activation in CB1 involve TM7, so its placement appears not to be as important as the other TMs. Also, none of the TM7 residues discussed in **Table 3.5** are conserved in TM7. Thus, overall, there is agreement between our predicted structural changes during activation and the experimentally observed ones.

Table 3.5 Comparison of the alterations in CB1 during activation with features seen in crystallized GPCRs. The observed structural changes between the constitutively active WT and the fully active T3.46I are similar to those which are observed in crystal structures. The cases where the CB1 models show deviations (red) from the crystal models are indicated in red.

Crystallized GPCR	Crystallized GPCR portion	CB1 model portion	Change during activation of crystallized GPCRs	Change from WT-3 to T3.46I-4
bRho/bOps (43), $HA_{2A}AR$ (53)	TMs 1 – 4	TMs 1 – 4	Not much change	Not much change (except TM4)
bRho/bOps, $HA_{2A}AR$	TMs 5 – 7	TMs 5 – 7	Large amount of movement	Large amount of movement
bRho/bOps	TM5	TM5 (M5.72)	Cytoplasmic end moves toward TM6 by 2 - 3 Å	Cytoplasmic end moves toward TM6 by 4.8 Å
bRho/bOps	TM6	TM6 (D6.25)	Cytoplasmic end moves outward by 6 – 7 Å	Cytoplasmic end moves outward by 6.6 Å
bRho/bOps	TM5 and TM6	TM5 and TM6	Become parallel to each other	Become parallel to each other
bRho/bOps	TM3 and TM6	TM3 and TM6	Broken ionic lock	Broken ionic lock
bRho/bOps	TM5 and TM6	TM5 and TM6	New salt bridge	New salt bridge
$H\beta_2AR$/nano-body (45)	E6.30	D6.30	Moves 11.4 Å	Moves 7.4 Å
$H\beta_2AR/G_S$ (41)	E6.30	D6.30	Moves 14.7 Å	Moves 7.4 Å
$H\beta_2AR$/nano-body	S5.46	V5.46	Moves inward by 2.1 Å	Moves inward by 2.2 Å
$H\beta_2AR/G_S$	S5.46	V5.46	Moves inward by 1.8 Å	Moves inward by 2.2 Å
$H\beta_2AR$/nano-body	P5.50	L5.50	Moves inward by 1.4 Å	Moves inward by 2.2 Å
$H\beta_2AR/G_S$	P5.50	L5.50	Moves inward by 1.6 Å	Moves inward by 2.2 Å
$HA_{2A}AR$	W6.48	W6.48	Moves inward by 1.9 Å	Moves inward by 1.3 Å
$HA_{2A}AR$	S7.42	C7.42	Moves closer to TM3 by 2 Å	Moves closer to TM2 by 3.5 Å
$HA_{2A}AR$	H7.43	L7.43	Moves closer to TM3 by 2 Å	Moves away from TM3 by 4.8 Å
$HA_{2A}AR$	N6.55	M6.55	Fixed	Moves by 0.9 Å
$HA_{2A}AR$	TM6	TM6 (D6.25)	Cytoplasmic end moves outward by 3 – 4 Å	Cytoplasmic end moves outward by 6.6 Å
$HA_{2A}AR$	NPxxY	N7.49	Moves inward by 4 – 5 Å	Moves inward by 3.5 Å
$HA_{2A}AR$	Y7.53	Y7.53	Moves by 5 Å, rotameric switch	Moves by 4.0 Å, no rotameric switch

We found the RMSD values of the alpha-carbons in the helical backbone of the constitutively active WT and double mutant receptors (T3.46A/R2.37A and T3.46A/R2.37Q) versus the highly constitutively active mutants (L3.43A and T3.46I) as seen in **Table 3.6A**. The range of RMSDs is 1.62 Å (the difference between T3.46A/R2.37A and T3.46I) to 3.08 Å (the difference between T3.46A/R2.37Q and L3.43A) with most of the values at approximately 2.5 Å. **Tables 3.6B-D** show the RMSD between antagonist or inverse agonist-bound and agonist-bound or putatively active conformations of crystallographically characterized GPCRs (10; 16; 18; 41; 42; 44; 53; 54). The CB1 WT receptor has RMSDs of 2.12 Å and 2.69Å with respect to T3.46I and L3.43A in that order. These values are in good agreement with those of the Hβ_2AR, 2.48 Å, and the boRho, 2.31 Å and 2.34 Å, as seen in **Tables 3.6B** and **C**. The RMSD values for T3.46A/R2.37A vs. L3.43A and T3.46A/R2.37Q vs. T3.46I are also in good agreement with the experimental results with RMSDs of 2.57 Å and 2.31 Å respectively. The RMSD between T3.46A/R2.37A and T3.43I, 1.62 Å, is a little lower than the other ones in **Table 3.6A**, but it matches well with the values from the HAA2AR, which are 1.61 Å and 1.67 Å. The RMSDs of the CB1 predicted structures show that there are significant differences between the constitutively active and highly constitutively active conformations, and these differences are in good agreement with those from experiments.

Table 3.6 RMSDs indicating change in helical packing upon activation. **(A)** Backbone RMSD comparison of the alpha-carbon TM regions backbone of the constitutively active (WT, T3.46A/R2.37A, and T3.46A/R2.37Q) and highly constitutively active (T3.46A and L3.43A) receptors. One structure was selected from the top ten low energy conformations of each receptor that best represents the consensus of the salt bridge and hydrogen bonds. The WT receptor is the third lowest energy conformation, T3.46A/R2.37A is the lowest, T3.46A/R2.37Q is the third lowest, and T3.46I and L3.43A are both the fourth lowest energy conformations. **(B)** Backbone RMSD comparison of crystallized inverse agonist carazolol-bound human β_2 adrenergic receptor (PDB ID: 2RH1) (18) with the agonist BI-167107 and G_s subunit-bound conformation (PDB ID: 3SN6) (41). **(C)** Backbone RMSD comparison of crystallized inverse agonist 11-*cis*-retinal-bound bovine rhodopsin (dark-state) (PDB ID: 1U19, chain A) (16) with the active 11-*trans*-retinal-bound metarhodopsin II (PDB ID: 3PQR) (44) and the putatively active ligand-free bovine opsin receptor (PDB ID: 3CAP, chain A) (42). **(D)** Backbone RMSD comparison of crystallized antagonist ZM241385-bound human adenosine A_{2A} receptor (PDB ID: 3EML) (10) with the active agonist UK-432097-bound conformation (PDB ID: 3QAK) (53) and the active agonist NECA-bound conformation (PDB ID: 2YDV) (54).

A.

CB1	**Wild-type**	**T3.46A/R2.37A**	**T3.46A/R2.37Q**
T3.46I	2.12	1.62	2.31
L3.43A	2.69	2.57	3.08

B

Human β_2 adrenergic receptor	**2RH1 (inverse agonist-bound)**
3SN6 (agonist and G_s-bound)	2.48

C.

Bovine rhodopsin receptor	**1U19 (inverse agonist-bound)**
3PQR (agonist-bound)	2.34
3CAP (ligand free, putatively active)	2.31

D.

Human adenosine A_{2A} receptor	**3EML (antagonist-bound)**
3QAK (agonist bound)	1.61
2YDV (agonist-bound)	1.67

The RMSD between the two highly constitutively active receptors, T3.46I and L3.43A, is 2.30 Å, which is a discernible difference. These dissimilar active conformations agree with the ensemble state theory, which suggests that each GPCR has multiple active conformations that are stabilized by different agonists or intracellular proteins. For example, cAMP assays indicate the activated WT CB1 receptor binds to the G_i protein (55), but the L3.43A mutant binds to the G_s protein (46). Different receptor conformations are potentially necessary for interacting with different types of G proteins. The predicted structures of T3.46I and L3.43A have different global conformations, so the T3.46I receptor may activate the G_i protein rather than the G_s one like the L3.43A mutant.

In addition to determining the backbone RMSD values, which may not be the best measure for comparing receptor conformations, we calculated the angle root-mean-square deviation (aRMSD). One of the major components of the GEnSeMBLE methodology is sampling the helix tilt, sweep, and rotation angles in order to find an ensemble of low energy conformation. **Table 3.7A** shows each model of the various receptors and their corresponding angles. Clearly, a single or double point mutation changes the entire packing of the helix bundle. As shown in **Table 3.7B**, the largest aRMSD values are 62.8° and 74.5°, which is the difference between the inactive T3.46A and active T3.46I and L3.43A receptors respectively. The difference between the constitutively active WT receptor and the two active mutants is not as large, with values of 29.1° and 32.2°. Receptors with the same level of constitutive activity tended to have the smallest aRMSDs. The WT receptor and T3.46A/R2.37A double mutant have an aRMSD of 9.1°. Even though these receptors only shared one of the intracellular salt bridges, they have the most similar helix orientations. A brief MD simulation is expected to relax the double mutant so that it would look more like the WT. The next smallest aRMSD is 22.1°, which is the difference between the active mutants L3.43A and T3.46I. Again, even though they have different salt bridge patterns, their helix orientations are comparable. Overall, our results show that receptors with similar levels of activation have similar conformations as proven by their helix orientations. Receptors with the largest discrepancies in activity had the greatest aRMSDs indicating the greatest structural changes that can be correlated with the observed differences in functional activity.

Table 3.7 Angle RMSDs indicating change in helical packing upon activation. **(A)** Tilt, sweep, and rotation angles of the consensus model representing each receptor. The angles are calculated with respect to the initial inact$T\beta_1AR$ (PDB ID: 2VT4) (11). One predicted structure was selected from the top ten low energy conformations of each receptor that best represents the consensus of the salt bridge and hydrogen bonds. The WT, T3.46A/R2.37A, T3.46A/R2.37Q, T3.46I, and L3.43A receptors use the same structures as in **Table 3.6**. The T3.46A receptor is represented by the lowest energy conformation, and L3.43A/D2.36A is represented by the seventh lowest energy conformation. **(B)** Angle RMSD (aRMSD) comparison of the different receptor consensus structures.

A.

Receptor	Tilt Angle							Sweep Angle							Rotation Angle						
	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7
T3.46A-1	10	0	-10	10	-10	0	0	-30	0	-30	0	15	30	0	-120	-150	-15	30	-150	15	-90
D2.63A/L3.43A-7	-10	-10	-10	10	0	0	0	-30	-30	30	0	0	15	-15	-30	-165	-15	30	-150	-30	-75
R2.37A/T3.46A-1	-10	-10	-10	10	0	-10	-10	-30	0	0	-30	-15	15	-15	90	-60	0	30	-150	15	-105
R2.37Q/T3.46A-3	-10	0	0	0	0	-10	-10	-15	30	0	-15	-15	0	-15	0	-60	0	30	-150	30	-60
WT-3	-10	-10	-10	10	0	0	0	-30	0	15	-30	-15	0	-15	105	-60	0	0	-150	15	-105
T3.46I-4	-10	-10	-10	10	0	-10	-10	-30	0	0	-15	0	-15	-15	120	-60	0	30	-90	-15	0
L3.43A-4	10	-10	-10	10	0	-10	-10	-30	0	15	-30	-30	0	-30	180	-60	-15	30	-135	-60	-15

B.

	T3.46A	D2.63A/L3.43A	R2.37A/T3.46A	R2.37Q/T3.46A	WT	T3.46I	L3.43A
T3.46A	0.0	27.9	51.9	37.2	55.9	62.8	74.5
D2.63A/L3.43A	27.9	0.0	38.8	46.6	41.3	47.0	55.1
R2.37A/T3.46A	51.9	38.8	0.0	24.1	9.2	29.1	33.5
R2.37Q/T3.46A	37.2	46.6	24.1	0.0	27.8	34.8	46.6
WT	55.9	41.3	9.2	27.8	0.0	29.1	32.2
T3.46I	62.8	47.0	29.1	34.8	29.1	0.0	22.1
L3.43A	74.5	55.1	33.5	46.6	32.2	22.1	0.0

Straight homology models cannot explain the functional effects of single-point mutations

By mutating a single residue, the receptors' respective global conformations change. In the T3.46I mutant, the larger hydrophobic isoleucine side chain replaces the polar threonine, breaks the hydrogen bond, and pushes the surrounding helices, including TM6, away in order to provide enough room to avoid side chain clashes. In the T3.46A mutant, the alanine residue is smaller than threonine and can fit in-between TMs 2 and 6. These two helices do not have to be as far away to avoid clashes as they would in the WT or T3.46I receptors. The T3.46A residue, however, does not interact with S2.45, so TM2 is rotated to optimize other intracellular polar interactions. In the L3.43A mutant, the threonine residue is left intact and thus so is the S2.45 and T3.46 hydrogen bond. The position of TM6 in relation to TM3 can change because the new alanine residue takes up less space than the original leucine one. The residues in the middle of TM6, such as V6.43, would clash with the 3.43 residue if the WT sequence had the L3.43A mutant conformation. In the L3.43A mutant, the extracellular end tilts towards TM3, but the intracellular end moves away breaking the R3.50+D6.30 interaction.

Figure 3.13 shows the results of the amino acid substitution without altering the helix packing (a homology model). **Figure 3.13A** shows the homology model of WT based on T3.46A, obtained by mutating the A3.46 residue to T3.46. The larger side chain of the T3.46 residue clashes with the TM2 backbone and would result in extremely unfavorable energies. Therefore, TM2 and/or TM3 must move to accommodate the T3.46 residue, affecting the packing of other five helices. Similarly, **Figure 3.13B** shows the homology model of T3.46I based on WT, obtained by mutating the T3.46 residue to I3.46. The larger isoleucine residue collides with S2.45. Thus, TM3 must shift to accommodate the presence of the larger side chain. However, it only moves slightly, sweeping away from TM2 by 15°. These helix movements cause large changes in the rest of the TM helix bundle in order to maximize favorable interactions, which cannot be captured by homology models even after relaxation by MD methods.

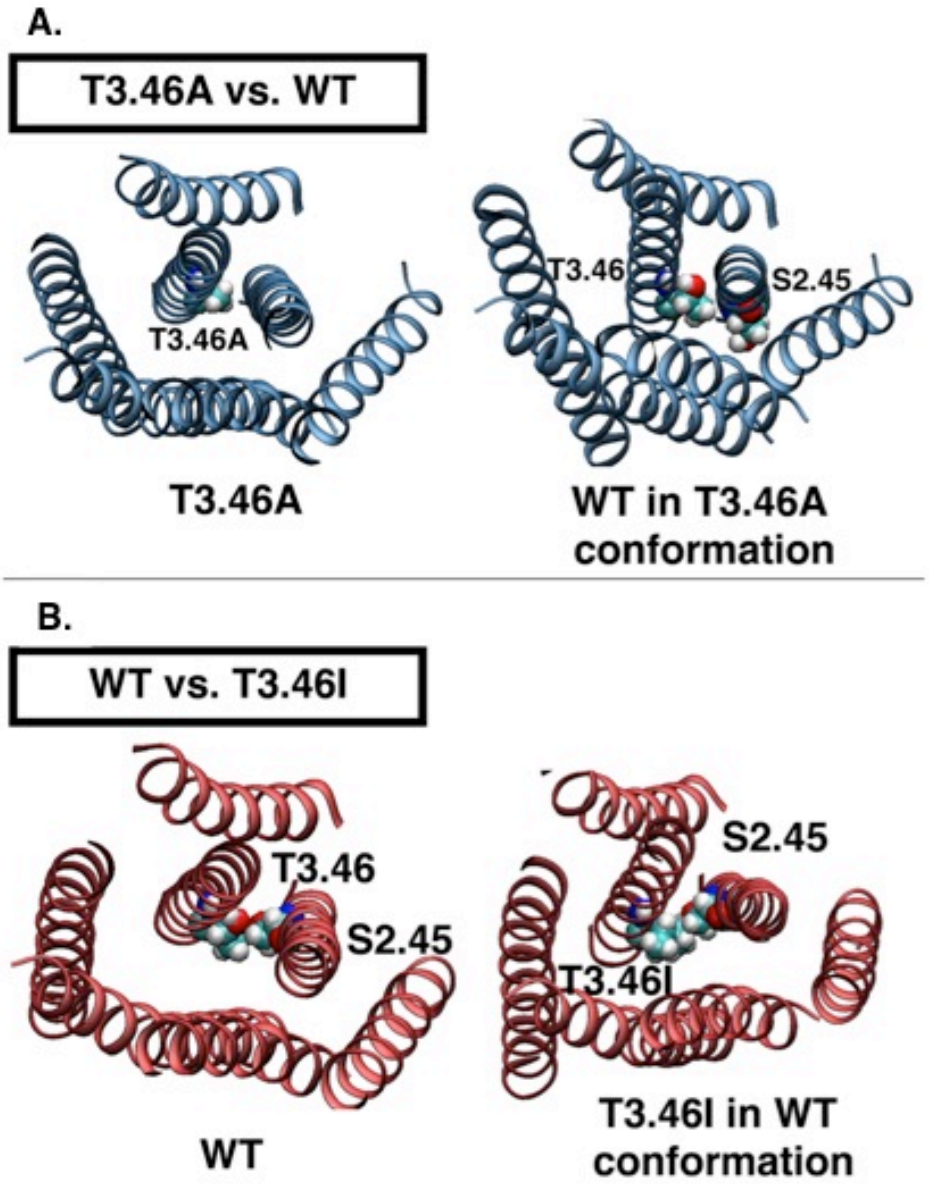


Figure 3.13 Effect of the amino acid at position 3.46 on the helical interactions in the human CB1 receptor. **(A)** Comparison of the T3.46A model versus the WT homology model (based on the T3.46A conformation). **(B)** Comparison of the WT model versus the T3.46I homology model (based on WT conformation). In each case, the residue (T in **(A)** or I in **(B)**) is too large for their respective positions, which explains why the receptor structures must change to accommodate the mutations.

Comparison to previous CB1 modeling studies

Previous CB1 modeling studies used homology modeling to bRho or other crystallized GPCRs as templates, followed in some cases by MD (20; 46; 56-60). Initial studies by Reggio and co-workers used the bRho template to develop homology models of CB1 (20; 56-58). Significantly kinked TM6 conformations (61) were added to the predicted 7-helix bundle, forming the R3.50+D6.30 that had been observed in the crystallographically characterized structure of bRho.

This predicted structure included another TM3+TM6 coupling between extracellular residues K3.28 and D6.58 (20). The CB1 active conformation was created by using activation hints from bRho: TM6 was straightened, and TMs 3 and 6 were rotated counter-clockwise with respect to the extracellular side. These motions broke both TM3+TM6 couplings (20).

A more recent model has used the inact$H\beta_2AR$ crystal structure as a homology template to provide secondary structure information. Only two salt bridges were observed—one between R3.50 and D6.30 as in the models discussed above and another between K3.28 and D2.63 (60) rather than D6.58 as proposed from the bRho homology structure (20). In general, these previous models agree that the conserved R3.50+D6.30 interaction is present in the constitutively active state. Shim's 2010 review (59) proposes that the activation mechanism CB1 consists of the agonist binding to EC2, which causes this loop region to couple to the aromatic microdomain in TMs 5 and 7, thereby forcing W6.48 to undergo a toggle switch. This movement, in turn, causes a cascade of interactions between hydrophobic or polar residues in the intracellular direction resulting in the breakage of the R3.50+D6.30 interaction (59). This mechanism has yet to be experimentally verified.

Software programs such as Modeller (62) have also been used previously to predict CB1's structure. The ModBase dataset (http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi) contains multiple CB1 structure predictions constructed from Modeller software (63). One of its more reliable conformations, according to its MPQS score, is based on the agonist-bound act$HA_{2A}AR$ crystal (PDB ID: 2YDV chain A) (64) homology model. In this predicted conformation R3.50 does not form an ionic lock with D6.30 even though both residues are conserved. As mentioned above, this particular interaction has been observed in multiple class A GPCRs. Furthermore in these homology-built structures, ligand-binding residues are not accessible to the binding site (20; 57). For example W5.43A leads to a ~1000 fold drop in binding affinity for the inverse agonist/antagonist rimonabant (57), but in the ModBase prediction, this residue faces into the lipid membrane region. In our predicted structures, however, this residue points into the binding site. Thus, the homology-based structures do not account for the essential characteristics of the CB1 receptor. None of the previous structures have implicated the TM2+TM6 coupling as being critical for the fully inactive CB1 receptor.

CONCLUSIONS

Using the CB1 receptor as the model, we provide the first complete picture of the GPCR activation mechanism from the inactive form (T3.46A) through the weakly constitutively active form (WT) to a highly constitutively active form (T3.46I). Some molecular signatures of our mechanism have been observed in crystal structures of inactive and active forms, while others are novel and may be specific to the cannabinoid family. Based on our understanding of the unique interaction observed between TMs 2 and 6 in the inactive T3.46A receptor, we designed double mutants expected to be constitutively active, and subsequently confirmed this with GTPγS assays. This lays the groundwork for predicting inactive and active mutations for this and other GPCRs, which may be used to determine new ligands to have specific activity for activating particular G proteins, β arrestins, and perhaps other functionalities.

The ability of our first principles-based ensemble methods to predict structures that explain experiments and to predict novel mutations with specific functionality, subsequently validated by experiments, suggests that these methods may be very useful in understanding activation and specific ligand-GPCR interactions for other systems.

Our experimental and computational studies show that dramatic changes in the global receptor conformations can be caused by the mutation of a single residue. This invalidates the implicit assumption in using mutation-binding studies to validate a model ligand-receptor structure, where it is assumed that the overall global structure of the receptor changes little. Our study shows cases in which this assumption is wildly invalid, requiring careful reinterpretation of all such GPCR binding-mutagenesis studies.

REFERENCES

1. Kenakin T, Miller LJ (2010) Seven transmembrane receptors as shapeshifting proteins: The impact of allosteric modulation and functional selectivity on new drug discovery. Pharmacol Rev **62**:265-304.
2. D'Antona A, Ahn KH, Kendall DA (2006) Mutations of CB_1 T210 produce active and inactive receptor forms: Correlations with ligand affinity, receptor stability, and cellular localization. Biochemistry **45**:5606-5617.
3. Ballesteros JA, Weinstein H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In: Sealfon SC, Ed. (1995) Receptor Molecular Biology. Academic Press, Waltham, Massachusetts, pp. 366-428.

4. Abrol R, Kim S-K, Bray JK, Griffith AR, Goddard WA, III (2011) Characterizing and predicting the functional and conformational diversity of seven-transmembrane proteins. Methods **55**:405-414.
5. Abrol R, Griffith AR, Bray JK, Goddard WA, III. Structure prediction of G protein-coupled receptors and their ensemble of functionally important conformations. In: Vaidehi N, Klein-Seetharaman J, Eds. (2012) Membrane Protein Structure: Methods and Protocols. Humana Press, New York, NY.
6. Abrol R, Bray JK, Goddard WA, III (2012) BiHelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. Proteins: Struct, Funct, Bioinf **80**:505-518.
7. Abrol R, Kim S-K, Bray JK, Trzaskowski B, Goddard WA, III (2013) Conformational ensemble view of G protein-coupled receptors and the effect of mutations and ligand binding. Methods Enzymol **520**:31-48.
8. Bray JK, Abrol R, Goddard WA, III, Trzaskowski B, Scott CE (2014) SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations. Proc Natl Acad Sci U S A **111**:E72-E78.
9. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res **25**:3389-3402.
10. Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman AP, Stevens RC (2008) The 2.6 Å crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. Science **322**:1211-1217.
11. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG, Schertler GFX (2008) Structure of a β_1-adrenergic G-protein-coupled receptor. Nature **454**:486-491.
12. Rasmussen SGF, Choi H-J, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VRP, Sanishvili R, Fischetti RF, Schertler GFX, Weis WI, Kobilka BK (2007) Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. Nature **450**:383-387.
13. Shimamura T, Hiraki K, Takahashi N, Hori T, Ago H, Masuda K, Takio K, Ishiguro M, Miyano M (2008) Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. J Biol Chem **283**:17753-17756.
14. Murakami M, Kouyama T (2008) Crystal structure of squid rhodopsin. Nature **453**:363-367.
15. Chien EYT, Liu W, Zhao QA, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, Stevens RC (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. Science **330**:1091-1095.
16. Okada T, Sugihara M, Bondar A-N, Elstner M, Entel P, Buss V (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. J Mol Biol **342**:571-583.
17. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science **330**:1066-1071.
18. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H-J, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human β_2-adrenergic G protein–coupled receptor. Science **318**:1258-1265.
19. Visiers I, Braunheim BB, Weinstein H (2000) Prokink: a protocol for numerical evaluation of helix distortions by proline. Protein Eng **13**:603-606.
20. Hurst DP, Lynch DL, Barnett-Norris J, Hyatt SM, Seltzman HH, Zhong M, Song ZH, Nie JJ, Lewis D, Reggio PH (2002) N-(Piperidin-1-yl)-5-(4-chlorophenyl)-1-(2,4-dichlorophenyl)-4-methyl-1H -pyrazole-3-carboxamide (SR141716A) interaction with

LYS 3.28(192) is crucial for its inverse agonism at the cannabinoid CB1 receptor. Mol Pharmacol **62**:1274-1287.
21. Kam VWT, Goddard WA (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. J Chem Theory Comput **4**:2160-2169.
22. Li J, Edwards PC, Burghammer M, Villa C, Schertler GFX (2004) Structure of bovine rhodopsin in a trigonal crystal form. J Mol Biol **343**:1409-1438.
23. Stenkamp RE (2008) Alternative models for two crystal structures of bovine rhodopsin. Acta Crystallogr D Biol Crystallogr **64**:902-904.
24. Standfuss J, Xie G, Edwards PC, Burghammer M, Oprian DD, Schertler GFX (2007) Crystal structure of a thermally stable rhodopsin mutant. J Mol Biol **372**:1179-1188.
25. Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y (2002) Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. Proc Natl Acad Sci U S A **99**:5982-5987.
26. Teller DC, Okada T, Behnke CA, Palczewski K, Stenkamp RE (2001) Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). Biochemistry **40**:7761-7772.
27. Okada T, Le Trong I, Fox BA, Behnke CA, Stenkamp RE, Palczewski K (2000) X-ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. J Struct Biol **130**:73-80.
28. Salom D, Lodowski DT, Stenkamp RE, Le Trong I, Golczak M, Jastrzebska B, Harris T, Ballesteros JA, Palczewski K (2006) Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. Proc Natl Acad Sci U S A **103**:16123-16128.
29. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. Science **289**:739-745.
30. Nakamichi H, Buss V, Okada T (2007) Photoisomerization mechanism of rhodopsin and 9-*cis*-rhodopsin revealed by x-ray crystallography. Biophys J **92**:L106-L108.
31. Standfuss J, Edwards PC, D'Antona A, Fransen M, Xie G, Oprian DD, Schertler GFX (2011) The structural basis of agonist-induced activation in constitutively active rhodopsin. Nature **471**:656-660.
32. Vanni S, Neri M, Tavernelli I, Rothlisberger U (2009) Observation of "ionic lock" formation in molecular dynamics simulations of wild-type β_1 and β_2 adrenergic receptors. Biochemistry **48**:4789-4797.
33. Dror RO, Arlow DH, Borhani DW, Jensen MO, Piana S, Shaw DE (2009) Identification of two distinct inactive conformations of the β_2-adrenergic receptor reconciles structural and biochemical observations. Proc Natl Acad Sci U S A **106**:4689-4694.
34. Lim KT, Brunett S, Iotov M, McClurg RB, Vaidehi N, Dasgupta S, Taylor S, Goddard WA (1997) Molecular dynamics for very large systems on massively parallel computers: The MPSim program. J Comput Chem **18**:501-521.
35. Goddard WA, Kim SK, Li YY, Trzaskowski B, Griffith AR, Abrol R (2010) Predicted 3D structures for adenosine receptors bound to ligands: Comparison to the crystal structure. J Struct Biol **170**:10-20.
36. Liu W, Chun E, Thompson AA, Chubukov P, Xu F, Katritch V, Han GW, Roth CB, Heitman LH, Ijzerman AP, Cherezov V, Stevens RC (2012) Structural basis for allosteric regulation of GPCRs by sodium ions. Science **337**:232-236.
37. Moukhametzianov R, Warne T, Edwards PC, Serrano-Vega MJ, Leslie AGW, Tate CG, Schertler GFX (2011) Two distinct conformations of helix 6 observed in antagonist-bound structures of a β_1-adrenergic receptor. Proc Natl Acad Sci U S A **108**:8228-8232.
38. Doré AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, Hurrell E, Bennett K, Congreve M, Magnani F, Tate CG, Weir M, Marshall FH (2011) Structure of the

adenosine A_{2A} receptor in complex with ZM241385 and the xanthines XAC and caffeine. Structure **19**:1283-1293.
39. Ballesteros JA, Jensen AD, Liapakis G, Rasmussen SGF, Shi L, Gether U, Javitch JA (2001) Activation of the β_2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. J Biol Chem **276**:29171-29177.
40. Yao XJ, Parnot C, Deupi X, Ratnala VRP, Swaminath G, Farrens D, Kobilka B (2006) Coupling ligand structure to specific conformational switches in the β_2-adrenoceptor. Nat Chem Biol **2**:417-422.
41. Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK (2011) Crystal structure of the β_2 adrenergic receptor-G_s protein complex. Nature **477**:549-555.
42. Park JH, Scheerer P, Hofmann KP, Choe H-W, Ernst OP (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature **454**:183-187.
43. Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP, Ernst OP (2008) Crystal structure of opsin in its G-protein-interacting conformation. Nature **455**:497-502.
44. Choe HW, Kim YJ, Park JH, Morizumi T, Pai EF, Krauss N, Hofmann KP, Scheerer P, Ernst OP (2011) Crystal structure of metarhodopsin II. Nature **471**:651-655.
45. Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK (2011) Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. Nature **469**:175-180.
46. D'Antona A, Ahn KH, Wang L, Mierke DF, Lucas-Lenard J, Kendall DA (2006) A cannabinoid receptor 1 mutation proximal to the DRY motif results in constitutive activity and reveals intramolecular interactions involved in receptor activation. Brain Res **1108**:1-11.
47. Song ZH, Bonner TI (1996) A lysine residue of the cannabinoid receptor is critical for receptor recognition by several agonists but not WIN55212-2. Mol Pharmacol **49**:891-896.
48. Chin CN, Lucas-Lenard J, Abadji V, Kendall DA (1998) Ligand binding and modulation of cyclic AMP levels depend on the chemical nature of residue 192 of the human cannabinoid receptor 1. J Neurochem **70**:366-373.
49. Rhee MH, Nevo I, Levy R, Vogel Z (2000) Role of the highly conserved Asp-Arg-Tyr motif in signal transduction of the CB_2 cannabinoid receptor. FEBS Lett **466**:300-304.
50. Feng WK, Song ZH (2003) Effects of D3.49A, R3.50A, and A6.34E mutations on ligand binding and activation of the cannabinoid-2 (CB2) receptor. Biochem Pharmacol **65**:1077-1085.
51. Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, Desale H, Clemons B, Cahalan SM, Schuerer SC, Sanna MG, Han GW, Kuhn P, Rosen H, Stevens RC (2012) Crystal structure of a lipid G protein-coupled receptor. Science **335**:851-855.
52. Crocker E, Eilers M, Ahuja S, Hornak V, Hirshfeld A, Sheves M, Smith SO (2006) Location of Trp265 in metarhodopsin II: Implications for the activation mechanism of the visual receptor rhodopsin. J Mol Biol **357**:163-172.
53. Xu F, Wu HX, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC (2011) Structure of an agonist-bound human A_{2A} adenosine receptor. Science **332**:322-327.
54. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG (2011) Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. Nature **474**:521-525.
55. Howlett AC, Fleming RM (1984) Cannabinoid inhibition of adenylate-cyclase. Pharmacology of the response in neuro-blastoma cell-membranes. Mol Pharmacol **26**:532-538.

56. Hurst D, Umejiego U, Lynch D, Seltzman H, Hyatt S, Roche M, McAllister S, Fleischer D, Kapur A, Abood M, Shi S, Jones J, Lewis D, Reggio P (2006) Biarylpyrazole inverse agonists at the cannabinoid CB1 receptor: Importance of the C-3 carboxamide oxygen/lysine3.28(192) interaction. J Med Chem **49**:5969-5987.
57. McAllister SD, Rizvi G, Anavi-Goffer S, Hurst DP, Barnett-Norris J, Lynch DL, Reggio PH, Abood ME (2003) An aromatic microdomain at the cannabinoid CB_1 receptor constitutes an agonist/inverse agonist binding region. J Med Chem **46**:5139-5152.
58. Singh R, Hurst DP, Barnett-Norris J, Lynch DL, Reggio PH, Guarnieri F (2002) Activation of the cannabinoid CB1 receptor may involve a W6.48/F3.36 rotamer toggle switch. J Pept Res **60**:357-370.
59. Shim J-Y (2010) Understanding functional residues of the cannabinoid CB1 receptor for drug discovery. Curr Top Med Chem **10**:779-798.
60. Shim J-Y (2009) Transmembrane helical domain of the cannabinoid CB1 receptor. Biophys J **96**:3251-3262.
61. Barnett-Norris J, Hurst DP, Buehner K, Ballesteros JA, Guarnieri F, Reggio PH (2002) Agonist alkyl tail interaction with cannabinoid CB1 receptor V6.43/I6.46 groove induces a helix 6 active conformation. Int J Quantum Chem **88**:76-86.
62. Fiser A, Sali A (2003) Modeller: Generation and refinement of homology-based protein structure models. Methods Enzymol **374**:461-491.
63. Pieper U, Webb BM, Barkan DT, Schneidman-Duhovny D, Schlessinger A, Braberg H, Yang Z, Meng EC, Pettersen EF, Huang CC, Datta RS, Sampathkumar P, MS M, Sjolander K, Ferrin TE, Burley SK, Sali A (2011) MODBASE, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res **39**:465-474.
64. Magnani F, Shibata Y, Serrano-Vega MJ, Tate CG (2008) Co-evolving stability and conformational homogeneity of the human adenosine A_{2A} receptor. Proc Natl Acad Sci U S A **105**:10744-10749.

Chapter IV

PREDICTED BINDING OF INVERSE AGONISTS TO THE CANNABINOID CB1 G-PROTEIN COUPLED RECEPTOR AND ITS USE IN DRUG DESIGN

ABSTRACT

The human cannabinoid type 1 (CB1) G-protein coupled receptor (GPCR) is a potential anti-obesity drug target, but the structure remains unknown. Using the Monte Carlo GEnSeMBLE method, we previously predicted the optimum ensemble of conformations for CB1, which we used to explain the observed changes in the receptors' G protein coupling activity upon single-point mutations. Our predictions were validated by predicting double and triple mutants that reversed activity either by gaining or decreasing function depending on the initial constitutive activity level. The mutations were subsequently validated experimentally. Here, we use these substantiated CB1 conformations to predict the binding sites and energies of a series of known CB1-selective inverse agonists. Our results agree with available site-directed mutagenesis data and predict new interactions that can be tested experimentally. In the lowest energy complex, the ligand is anchored via hydrogen bonds with K3.28 and a tryptophan on the TM5 (W5.43), which is different from previously published pharmacophores. Seven derivatives of rimonabant were docked to CB1, and the calculated binding energies were compared with the experimental binding affinity resulting in a ~90% correlation. This excellent agreement between the calculations and experiments gives credence to our predicted CB1 structures and inverse agonist binding site. Using this binding pose, rimonabant was reimagined so that it could exploit interactions with residues in the binding site. Based on our predicted rimonabant-CB1 pharmacophore, we propose new anti-obesity drug candidates that are expected to be strong CB1-targeting inverse agonists that can be tested by experimentalists.

INTRODUCTION

G-protein coupled receptors (GPCRs) are located in the cellular membrane and act as mediators for cell signaling making them popular targets for medication, but drug design for GPCRs has been challenging due to the paucity of structures. These integral membrane proteins are extremely difficult to crystallize—only ~26 GPCRs have been done so successfully. Despite recent advances in structure determination methods (1), little is known about GPCR structures, especially their activation mechanisms. Only six GPCRs, bovine rhodopsin (bRho) (2-5), human β_2 adrenergic receptor ($H\beta_2AR$) (6; 7), human adenosine A_{2A} receptor (8-11), human muscarinic M2 receptor (12), human neurotensin NTS1 receptor (13), and human P2Y12 receptor (14) have been crystallized in the active or partially active conformations, of which only the $H\beta_2AR$ has been co-crystallized with the full heterotrimeric G_s protein (6).

The human cannabinoid 1 (CB1) receptor is an example of one of the many GPCRs that have not been crystallized, and therefore its spectrum of conformations, from a fully inactive state to various active states, remain unknown. CB1 has high therapeutic potential as a drug target. This receptor is activated by marijuana and its synthetic versions that had been developed to alleviate pain and increase appetite for AIDS and cancer patients. The CB1-selective inverse agonist/antagonist rimonabant (15) is an anti-obesity drug that was available in Europe and was in phase III of US FDA clinical trials, but had to be withdrawn from the European market due to serious side effects of depression and suicide. By determining a realistic binding site, we can develop drugs specifically for CB1 and decrease the probability of off-target interactions and subsequent unfortunate side effects.

In previous studies, we used the Monte Carlo GEnSeMBLE (GPCR Ensemble of Structures in Membrane Bilayer Environment) method to predict an ensemble of low energy CB1 conformations (16; 17). Briefly, the GEnSeMBLE method consists of sampling the energies of trillions of seven helix bundles by rotating and tilting the transmembrane (TM) helices. This method is unique for its ability to predict multiple conformations, unlike homology modeling. It is important to generate multiple conformations because receptors are dynamic and flexible. A single GPCR can bind to multiple types of intracellular proteins suggesting that it may undergo conformational changes to do so. Furthermore, different ligands bind to and stabilize different GPCR conformations (18-20). Previously, we used GEnSeMBLE to identify ten low energy CB1 conformations that are predicted to be energetically accessible. This current study focuses on validating our proposed CB1 receptor conformations by docking a known inverse agonist to them and analyzing our docked poses in light of the site-directed mutagenesis data (21-30) and structure-activity relationship (SAR) data (31; 32).

Site-directed mutagenesis data of residues around a ligand binding site indicate which ones have a large impact on a particular ligand's binding affinity and can perhaps identify ones that are in the binding site and interact with the ligand. If the mutated residues are located in the ligand's binding site, then we can postulate that the changes in ligand binding affinity are a result of the changes in the mutated residues' interactions with the ligand. A correlation between large changes in binding affinity and strong interactions with the ligand is expected. **Table 4.1** shows the side-directed mutagenesis data for rimonabant (**Figure 4.1**). Binding affinities are measured by K_i and K_d values. The mutations that have a significant impact on binding affinity to

rimonabant are W5.43A, F3.36A, and K3.28A, which cause >1000-fold (21), 20-fold (22), and 17-fold (23) decreases respectively.

Table 4.1 Compilation of site-directed mutagenesis data for CB1 residues.

Ligand	Mutation	TM#	Initial K_i or K_d[a](nM)	New K_i or K_d[a] (nM)	Fold Change
Rimonabant	***W5.43A***[c]	*5*	*4.8*	*46% at 5mm*	*>1000*[b]
Rimonabant	***F3.36A***[d]	*3*	*3.7*	*75.5*	*20.4*
Rimonabant	***K3.28A***[e]	*3*	*2.3*[a]	*39.6*[a]	*17.2*
Rimonabant	**I2.62T-D2.63N**[j]	2	3.3	22.7	6.9
Rimonabant	**W6.48A**[c]	6	4.8	33	6.9
Rimonabant	**D2.63N**[j]	2	3.3	17.6	5.3
Rimonabant	**C7.42M**[k]	7	39	194	5.0
Rimonabant	**F2.61A**[l]	2	2.3[a]	11.0[a]	4.8
Rimonabant	**F7.35A**[l]	7	2.3[a]	8.4[a]	3.6
Rimonabant	**S2.60A**[i]	2	7.2	22.8	3.2
Rimonabant	**L3.29A**[l]	3	2.3[a]	7.2[a]	3.1
Rimonabant	**M6.55A**[l]	5	2.3[a]	6.8[a]	3.0
Rimonabant	**F3.36A**[c]	3	4.8	14	2.9
Rimonabant	**T3.33A**[l]	3	2.3[a]	6.5[a]	2.8
Rimonabant	**I2.62T**[j]	2	3.3	8.8	2.7
Rimonabant	**F3.36L**[d]	3	3.7	7.3	2.0
Rimonabant	**F3.25A**[c]	3	4.8	9.6	2.0
Rimonabant	**S7.39A**[i]	7	18.4	14.2	0.77

[a] K_d value. [b] Italicized rows indicate large changes (greater than 10-fold) in binding affinity upon mutation. [c] See Ref.(21). [d] See Ref.(22). [e] See Ref.(23). [f] See Ref.(24). [g] See Ref.(25). [h] See Ref.(26). [i] See Ref.(27). [j] See Ref.(28). [k] See Ref.(29). [l] See Ref.(30).

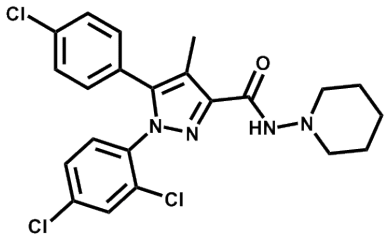


Figure 4.1 Schematic illustrations of the CB1-selective inverse agonist and anti-obesity drug rimonabant.

In order to validate our predicted CB1 receptor conformations, we docked rimonabant with the hierarchical programs, DarwinDock and GenDock (33-37), developed within our lab, so that we might determine a correlation between our calculated residue interactions and the change in binding affinity upon mutation. Our predicted ligand-receptor complexes can accurately explain site-directed mutagenesis data and, for the most part, the experimental trends in binding affinity.

These predicted binding sites were used to design and identify new ligands that we believe would act as inverse agonists and strongly interact with the CB1 receptor. In this study, we propose a binding site for rimonabant that has differences from what has been previously published (21; 23; 30; 38-40). The docked rimonabant pose serves as an inspiration for recommending potential CB1-selective inverse agonists.

METHODS

(For more details about ligand docking methodology, see *Chapter II*.)

We predicted an ensemble of CB1 receptor structures based on first principles using the GEnSeMBLE method developed within the lab (34; 41-44). Details of the procedure for structure prediction have been published previously (16; 17; 44). The procedure for docking rimonabant and other inverse agonists to the CB1 receptors are outlined below. All the simulations used the Dreiding III force fields (FF) (45) [with exponential-6 (exp6) or Lennard-Jones (LJ) forms for the van der Waals (vdW) term], unless otherwise noted.

Ligand preparation

For rimonabant, we constructed each ligand with Maestro software (46) and performed a conformational search with MacroModel software (47). A total of 83 ligand conformations were saved for rimonabant. Nine conformations remained after two rounds of clustering. One conformation was constructed from existing coordinates from the crystallized molecule in methanol solvate (48). Each ligand was minimized with SGB solvation (49) for 100 steps or to a convergence threshold of 0.2 kcal/mol/Å RMS force with MPSim software (50).

Docking of rimonabant to CB1 receptors

We docked ten conformations (nine from conformational search and one from a crystal) of rimonabant to the selected putative binding regions of all ten low energy structures for the CB1 receptor using the programs DarwinDock and GenDock (33-37).

DarwinDock

This program combined the completeness of generating poses for the binding pocket, indicated by the sphere regions, and the clustering of similar poses together to reduce the computational cost of energy scoring for the possible candidates. Darwin Dock consisted of using Dock6 (51) to

generate 5000 ligand poses in the alanized receptor for multiple steps. The W5.43 residue is the only hydrophobic, bulky residue not replaced with alanine because this tryptophan is deemed critical for ligand interaction by site-directed mutagenesis data (see **Table 4.1**), so it remains unchanged during docking. A ligand pose was deemed acceptable if it clashed or bumped the receptor residues at six positions or less. The complexes are clustered so that every family member was within a 2.00 Å RMSD of another. These steps are repeated until the number of new families created is less than 2% of the number of existing families. The program scores the energies of the family heads and selects the 10% of family heads with the lowest energies. All the members of these respective families are scored energetically. One hundred and fifty receptor-ligand poses are chosen according to the lowest phobic, polar, and total energy for enrichment in GenDock.

GenDock

GenDock refined and enriched the 150 docked ligand-receptor poses generated by DarwinDock. In the SCREAM module (45) alanine residues were replaced with the original hydrophobic residues, and the side chain rotamers of binding site residues were sampled before the entire complex was minimized for 10 steps. In the next module, the receptor was neutralized and minimized for 60 steps using the exp6 vdW term in the Dreiding FF. A final complex was selected for each receptor and ligand conformation (300 complexes total) according to the lowest binding energy with strain and ligand solvation, which we believe best represents the binding affinity. The binding energy with strain and ligand solvation is defined as the energy difference between the complex and the sum of the receptor and ligand energies with ligand strain and ligand solvation added in. The selected complex was minimized with the Dreiding III FF with the LJ vdW term (Dreiding III-LJ FF) in vacuum for 50 steps or to an RMS force threshold of 0.5 kcal/mol/Å using MPSim software (50).

Validation of predicting binding sites

Building and docking rimonabant derivatives for SAR studies

We took the lowest energy CB1 receptor-rimonabant complex shown in **Figure 4.2**, extracted the rimonabant ligand, modified it with Maestro software to look like the respective derivative, calculated the charges with Jaguar software, minimized it with for 100 steps or a threshold of 0.2 kcal/mol/Å with MPSim, and docked it using the same procedure described above with a few exceptions. Instead of docking to ten receptor conformations, we docked to one—the alanized CB1 receptor conformation WT6 because that produced the lowest energy complex with

rimonabant shown in **Figure 4.2**. We also docked a single ligand conformation derived from the docked rimonabant pose. We repeated these steps for seven derivatives and the original rimonabant ligand. For rimonabant, we did not alter the ligand structure and only docked the optimized ligand conformation. The final complexes were chosen according to the lowest binding energy with strain and ligand solvation. For comparison of the experimental binding affinities with the computational energies, we used the binding energy, or the difference in energy between the receptor-ligand complex and the receptor and the ligand structures separately. The ligand strain and ligand solvation were ignored since we are comparing energetics across different ligands.

Free energy and binding affinity calculations

The experimental change in free energy upon ligand binding to the receptor, ΔG_{bind}^{Exp}, is obtained from the pK_i (52) in **Equation 4.1**:

$$\Delta G_{bind}^{Exp} = -RT \ln\left(10^{-pK_i}\right) \qquad \textbf{(4.1)}$$

The $\Delta\Delta G^{\exp}$, or the difference in the change in experimental free energy upon ligand binding for a given inverse agonist (ΔG_{bind}^{Exp}) with respect to the corresponding value of rimonabant, is determined by **Equation 4.2**:

$$\Delta\Delta G^{\exp} = -RT \ln \frac{K_i\left(\text{Inverse agonist}\right)}{K_i\left(\text{Rimonabant}\right)} \qquad \textbf{(4.2)}$$

Discovery of novel CB1 ligands

Exploiting the ligand binding site to define a new scaffold

Our predicted rimonabant-CB1 binding site was used to identify new potential CB1-selective inverse agonists. We performed a crude search in PubChem (53) looking for ligands that are comparable to rimonabant according to the Tanimoto coefficient (54), which indicates how similar the 2D structures of the ligands are to one another. The Tanimoto coefficient ranges from 0 to 1, with 0 being no resemblance between the molecules and 1 being identical resemblance between the molecules. We wanted ligands that are modifications of rimonabant to exploit interactions with underused polar and aromatic residues in the predicted binding site. For example, K7.32 is located near rimonabant, but does not form a hydrogen bond with it, so we searched for a ligand with an appropriately placed functional group to create a desired polar interaction. We also wanted a ligand that is commercially available, so that experimentalists can

easily test our predictions, and one that has not been previously tested with the CB1 receptor. Once we had identified a ligand (**MSC1**, see **Figure 4.6**) that met the above criteria, we performed a search in PubChem to identify ligands that scored a 0.90 on the Tanimoto coefficient, or those that have a 2D structure that is 90% similar to the new PubChem ligand **MSC1**. The ligands found in PubChem are denoted by the acronym 'MSC' followed by a number (**MSC1** through **MSC16**).

RESULTS AND DISCUSSION

Docking inverse agonist rimonabant to CB1

The lowest energy pose for the CB1-rimonabant complex agrees with site-directed mutagenesis data because the ligand is anchored by hydrogen bonds with W5.43 and K3.28 (**Figure 4.2**). These residues have the largest impact by decreasing experimental binding affinity upon mutation as seen in **Table 4.1**. The breakdown of calculated binding energies for the complex is shown in **Table 4.2**. W5.43 has the largest non-bonding energy of all the residues in the receptor with -9.36 kcal/mol. As stated above, **Table 4.1** shows that the W5.43A mutation has a major effect on rimonabant binding. The experimental mutation indicates a significant decrease in binding affinity of >1000-fold. Thus, our docking studies agree with experimental data because rimonabant has significant polar and phobic contacts with the CB1 receptor.

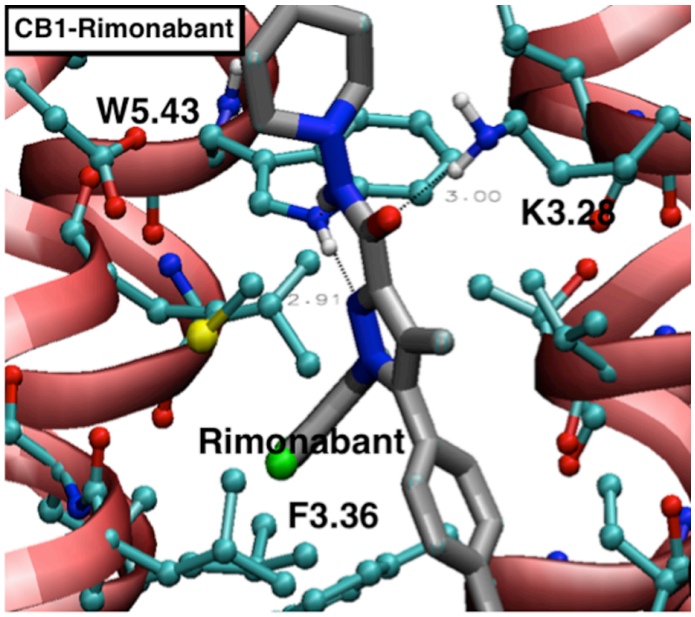


Figure 4.2 Inverse agonist rimonabant docked to CB1 receptor. Rimonabant is anchored to CB1 by the two hydrogen bonds: one with W5.43 and the other hydrogen bond with K3.28. Hydrogen bond heteroatom distances are labeled in black.

Table 4.2 Cavity analyses for the CB1 WT in complex with rimonabant.

Residue[a]	Number[b]	VdW[c]	Coulomb[c]	H-Bond[c]	NonBond[c,d]
TRP[e]	**5.43**	**-3.146**	**-2.085**	**-4.129**	**-9.360**
MET	6.55	-6.043	-0.756	0	-6.799
LYS	**3.28**	**-0.833**	**-1.410**	**-3.322**	**-5.566**
PHE	3.36	-3.841	-0.054	0	-3.896
VAL	3.32	-2.904	-0.363	0	-3.267
THR	5.47	-1.912	-0.048	0	-1.960
LEU	6.51	-2.100	0.193	0	-1.907
TRP	6.48	-1.844	0.008	0	-1.837
THR	3.33	-1.889	0.093	0	-1.797
ILE	2.56	-1.449	0.000	0	-1.449
LEU	7.43	-1.382	0.119	0	-1.263
PHE	7.35	-0.959	-0.174	0	-1.134
SER	3.35	-1.385	0.302	0	-1.083

[a] Only residues with energies whose absolute values are greater than 1.0 kcal/mol.
[b] Residues are numbered according to the Ballesteros-Weinstein scheme (55).
[c] All energies are in units of kcal/mol.
[d] The non-bonding (NonBond) energy in the far right column is the sum of the van der Waals (VdW) energy, coulombic (Coulomb) energy, and hydrogen bond (H-bond) energy.
[e] Residues with hydrogen bonds are highlighted in bold red.

According to our docking cavity analysis in **Table 4.2**, K3.28 has the third largest non-bonding interaction energy, -5.566 kcal/mol, with rimonabant due to its hydrogen bond with the amide carbonyl of the ligand. This docking pose is supported by a previous SAR study (23), which showed that the binding affinity decreased by 17-fold by mutating K3.28 to alanine. However, the same experiments with the rimonabant derivative VCHSR, a version of rimonabant with hydrocarbons replacing the amide and piperdine groups, showed its the binding affinity relative to rimonabant did not change for the K3.28A mutant. This study suggests that K3.28 interacts with the polar atoms of the amide group or piperdine ring, including the carbonyl (23), which agrees with the predicted docked ligand pose.

In our docked pose of rimonabant with the WT receptor, the F3.36 residue has a significant interaction with the ligand consisting of a van der Waals interaction energy of -3.896 kcal/mol. In binding assays, the F3.36A mutation decreased this inverse agonist's binding affinity by 20-fold (22), or 1.78 kcal/mol, for the human CB1 receptor. An earlier study done with the mouse sequence showed that F3.36A only decreases the binding affinity by 3-fold (21), which is not a significant amount (**Table 4.1**). Returning to the human sequence studies, the F3.36L mutation

decreased the binding affinity by only 2-fold indicating the importance of the bulky hydrophobic residue for receptor-ligand interaction (22).

We notice that M6.55 has a strong interaction with rimonabant, but that is not explained by the site-directed mutagenesis data. This residue has the second strongest interaction with CB1 with a non-bonding energy of -6.8 kcal/mol, which is almost entirely hydrophobic in nature (**Table 4.2**). When methionine is mutated to an alanine, however, the binding affinity only decreases slightly, by 3-fold (30), or by 0.65 kcal/mol in energy, for the CB1 receptor. The predicted binding site cannot explain this site-directed mutation.

Our predicted rimonabant binding site in CB1 is different from previous computational studies that examined rimonabant docked to bRho templates of human CB1 (21; 23; 39; 40). **Figure 4.3** compares this rimonabant pharmacophore published by Lange and Kruse (38) (**Figure 4.3A**) with the corresponding pharmacophore from this study (**Figure 4.3B**). Our models are basically perpendicular to one another. Rimonabant in the previous models (**Figure 4.3A**) spans the width of the binding site with the chlorophenyl rings near TM5 and the piperdine ring near TM3. In **Figure 4.3B**, the ligand is lying parallel to the *z*-axis passing from the extracellular to intracellular side of the membrane. Both agree that a hydrogen bond forms between K3.28 and the amide carbonyl of rimonabant. However, the earlier models show that K3.28 also participates in a salt bridge with D6.58 (21; 23; 39), which we do not observe. In addition, in previous models, the W5.43 residue is sandwiched by rimonabant's two chlorophenyl groups and thus has strong aromatic interactions. Furthermore, multiple other aromatic residues in the binding pocket (F3.25, F3.36, W4.64, Y5.39, W6.48) participate in stacking with the ligand (21; 39; 40), whereas in our predicted binding site, the W5.43 residue's impact on binding affinity comes from its ability to form a hydrogen bond with rimonabant. Instead of W5.43, the F3.36 residue is sandwiched by the chlorophenyl groups.

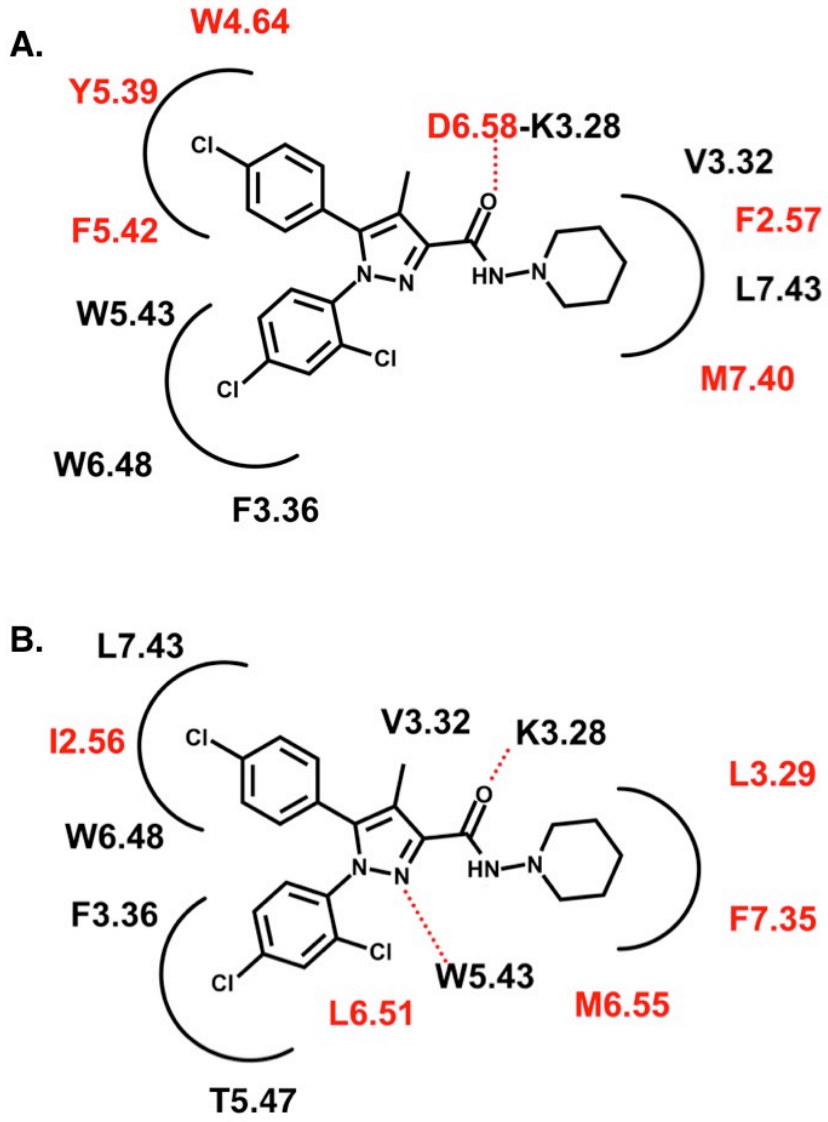


Figure 4.3 CB1-rimonabant pharmacophores with side chains important for receptor-ligand interactions labeled. **(A)** Lange and Kruse published a pharmacophore (38) based on prior docking studies (21; 23; 39; 40) involving the human CB1 receptor predicted using homology modeling with bRho. These studies find that rimonabant forms one hydrogen bond with K3.28 and sandwiches W5.43. [**Figure 4.3A** is based on the previously published **Figure 4** from Lange and Kruse, 2005 (38).] **(B)** Our predicted pharmacophore shows that rimonabant has a different binding site in our GEnSeMBLE-derived CB1 structure that allows it to form two hydrogen bonds, including one with W5.43, and have strong aromatic interactions with the receptor. Residues in the binding site are labeled in both **Figures 4.3A** and **4.3B**. Residues labeled in red in **Figure 4.3B** are unique to that particular pharmacophore and are not found in the binding site shown in **Figure 4.3A**.

A recent paper from Shim *et al.* based its CB1 model on the Hβ_2AR (30), and the authors identified a similar binding site compared to the one discussed above that was published by Lange and Krause (38). The ligand does lie perpendicularly to the model presented in this study, in that the piperdine ring points to the TM1-2-7 binding pocket and the chlorophenyl rings point

towards TM5 where they form aromatic stacks with W5.43 and W6.48. However, rimonabant's binding site is markedly different from the others because this particular ligand pose does not form a direct hydrogen bond with K3.28. Also, the monochlorophenyl and dichlorophenyl rings are switched compared to the previous models.

It is reasonable for our predicted rimonabant binding site to disagree with the previously published ones since our predicted receptor structures are different. Details on prediction methods and a structural comparison have been published previously (16; 17; 44), but in general, we based our structure on the turkey β_1 adrenergic receptor and performed a 360° rotation of all the helices and additionally sampled helix tilt and sweep angles. To reiterate, the previous studies used homology models based on the bRho or Hβ_2AR structure templates. These three GPCRs have very similar global conformations, but we preformed considerable sampling to obtain CB1 receptor conformations that were significantly different than the starting template. The method we used to identify our final receptor-inverse agonist complex is also different. The most recent study from Shim *et al.,* (30) used the selection criterion that one of the two chlorophenyl rings should be near the TM3-5-6 hydrophobic site, so their final pose was chosen based on rimonabant's interactions with residues for which there was site-directed mutagenesis data, whereas we selected our complex based on the lowest energy complex, which agrees with experimental data.

The conformation that produced a rimonabant binding pose most similar to ours came from another study where the ligand was docked to a bRho homology model (56). This predicted binding site is similar to the other bRho-based models mentioned above in that it has a K3.28+D6.58 salt bridge, but in contrast with those studies and in agreement with ours, rimonabant sits vertically in the binding pocket—the piperdine ring points towards the extracellular end and the chlorophenyl rings point towards the intracellular end. The oxygen atom of rimonabant's amide carbonyl forms a hydrogen bond with K3.28, but unlike ours, the ligand does not form any hydrogen bonds with W5.43. The amide carbonyl points toward TM5, while the amide proton points to the TM1-2-7 pocket. They predict that the pyrazole nitrogen on rimonabant forms a hydrogen bond with Y5.39, which we do not observe in our complexes.

Analysis of the conformational ensemble of CB1 structures

Our structure prediction methodology is unique in that we can predict an ensemble of energetically accessible conformations. This procedure is a more realistic approach than

homology modeling because the receptors are flexible and dynamic, and experimental evidence suggests that different ligands can stabilize different receptor conformations that can potentially interact with various intracellular proteins. Surprisingly, rimonabant does not have its most favorable binding energies with the lowest energy CB1 receptor conformation (called WT1) (**Table 4.3**). Instead, rimonabant has a slightly better binding affinity with the conformation ranked sixth in energy (WT6).

Table 4.3 Comparison of receptor conformations selected by the inverse agonist rimonabant.

Energy Rank	WT Conf. # bound to Rimonabant[a]	Binding Energy with strain and Rimonabant solvation
1	**WT6**[b]	-59.10
2	**WT1**	-57.24
3	WT9	-55.44
4	**WT5**	-54.39
5	WT2	-50.97
6	WT7	-50.12
7	WT4	-50.05
8	WT10	-49.59
9	WT8	-49.35
10	WT3	-48.63

[a] The complexes are ranked according to best binding energy with strain and ligand solvation included.
[b] Receptor conformation numbers that do not contain the conserved R3.50 and D6.30 salt bridge are in bold red.

Even more surprising is that the inverse agonist rimonabant preferentially binds to receptor conformations that lack the R3.50 and D6.30 ionic lock, which is believed to be important for preventing activation (**Table 4.3**). Experiments have shown that rimonabant inactivates the CB1 receptor (15; 32), but the inverse agonist has the best binding energies with those that look activated. When comparing the lowest energy complex for each receptor conformation, the three conformations that lack the conserved R3.50 and D6.30 ionic lock are ranked in the top four rimonabant-CB1 complexes, which is not what we had anticipated.

There have been, on the other hand, several studies that suggest this R3.50 and D6.30 salt bridge is not necessary for maintaining the inactive conformation. It is possible for antagonists to bind to the inactive state without the ionic lock (57; 58). For example, Audet and Bouvier have shown that the R3.50 and D6.30 salt bridge is broken in the crystallized complex of Hβ_2AR bound to the antagonist carazolol. They also note that as of 2012, of the 36 crystallized GPCR structures bound to antagonists and inverse agonists, only three contain this intact salt bridge, which was not observed across all structures for these particular receptors (57). In previous work with the

human CB1 receptor, we had shown that a different residue, R2.37, plays an important role in preventing receptor activation (16; 17; 44). We believe that this residue forms an 'ionic lock' via a salt bridge interaction with D6.30 to stabilize the inactive conformation. Thus, the traditional R3.50 and D6.30 salt bridge seems unnecessary for stabilizing the inactive CB1 conformation, and the inverse agonist would not preferentially bind to a conformation that contained this signature contact. We did not dock rimonabant to the CB1 WT conformation containing the R2.37 and D6.30 salt bridge that we believe to be the ionic lock for this receptor because none of the low energy WT conformations contain this contact.

In its current state, the rimonabant-bound CB1 receptor conformation WT6 does resemble the predicted structure of a constitutively active mutant, T3.46A/R2.37A, which too lacks the R3.50 and D6.30 ionic lock (16; 17). Their alpha-carbon backbone RMSD is 1.2 Å, which is smaller than all crystal structure resolutions (11) and indicates that the two conformations are very similar. For comparison, the difference between the WT receptor conformations WT2 (which contains the R3.50+D6.30 salt bridge) and WT6 (which does not) is 1.8 Å.

Analysis of structure-activity relationship (SAR) with rimonabant derivatives

In order to validate our proposed binding site, we docked known derivatives of rimonabant (**Figure 4.4A**) to the CB1 receptor conformation and plotted our calculated binding energies against the experimental binding affinity, pK_i, to find the correlation between the two data (**Figure 4.5**). In all of the complexes, the ligand bound to the same site as rimonabant originally did (**Figure 4.4B**). **Figure 4.5A** shows a comparison between the pK_i of the inverse agonists from Hurst *et al.* (31) and our calculated binding energies. We observe a very high correlation of 93.4%, which shows there is good agreement between the trends observed in experimental binding affinities and our own calculated binding energies.

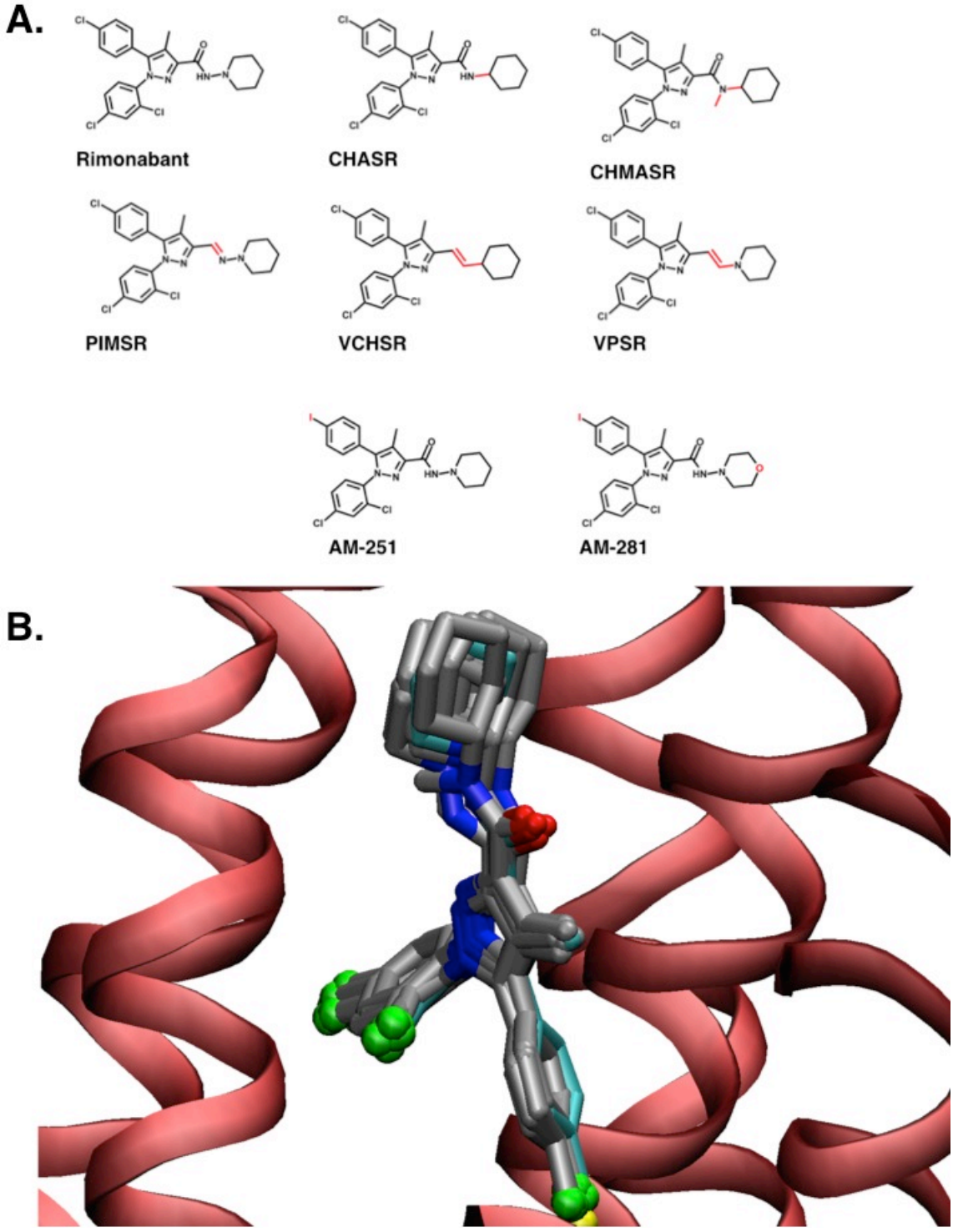
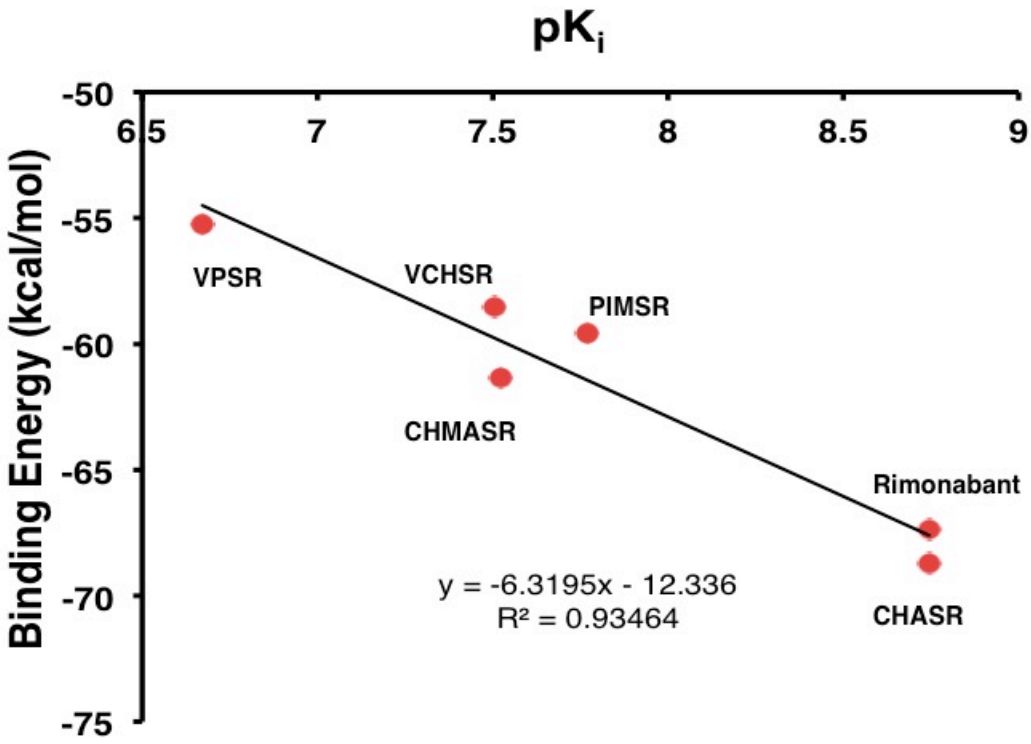


Figure 4.4 Rimonabant derivatives used in SAR study. **(A)** Rimonabant and the seven derivatives used in the SAR study. The portions of the ligand that are different from rimonabant are colored in red. The derivatives in the first two rows were used in the Hurst *et al.,* 2006 study (31), and the derivatives in the third row were used in the D'Antona *et al.,* 2006 studies (32; 59). **(B)** Rimonabant and seven derivatives docked to the CB1 complex. Rimonabant has carbon atoms colored in cyan. The other derivatives have carbon atoms colored in grey.

A.


pK$_i$
Binding Energy (kcal/mol)
-50
-55
-60
-65
-70
-75
6.5
7
7.5
8
8.5
9
VPSR
VCHSR
PIMSR
CHMASR
Rimonabant
CHASR
y = -6.3195x - 12.336
R² = 0.93464

B.

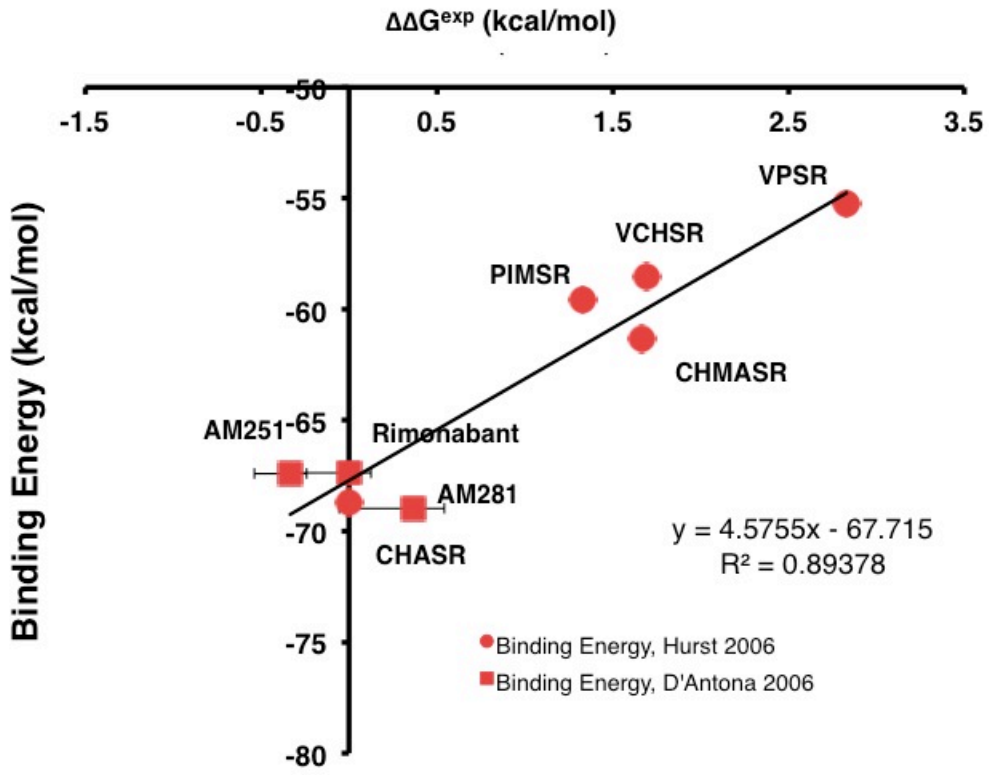


Figure 4.5 Comparison of experimental binding affinities and computationally calculated energies between receptor and inverse agonists. **(A)** Plot showing the correlation of the experimental binding affinity, pK_i, for inverse agonists in Hurst *et al.* (31) versus our calculated binding energy (red circles). R^2 values are indicated on the plot. **(B)** Plot showing the correlation of the $\Delta\Delta G^{exp}$ of previously published inverse agonists with respect to that of rimonabant (31; 32) versus the calculated binding energy. R^2 values are indicated on the plot.

For the inverse agonists from D'Antona *et al.* (32; 59), we compared these calculated energies with the ΔG_{bind}^{Exp} determined from **Equation 4.1**. The ΔG_{bind}^{Exp} values are within 1 kcal/mol of each other since the energies are -9.33 kcal/mol for AM-281, -10.0 kcal/mol for AM-251, and -9.70 kcal/mol for rimonabant. Similarly, both the calculated cavity and binding energies for the three ligands are approximately within 1 kcal/mol. The binding energies are -67.4 to -69.0 kcal/mol, which are very similar and agree with the experimentally observed negligible change in free energy upon binding. This excellent correlation is surprising because the predicted energies are based purely on enthalpy and do not include entropy. However, the entropy change upon binding of all the ligands should be similar since they all have the same number of rotatable bonds.

Figure 4.5B combines the results from **Figure 4.5A** with those regarding the inverse agonists from D'Antona *et al.* (32) for a single comparison of our calculated binding energies and the binding affinities from the two experiments. The respective $\Delta\Delta G^{exp}$ values from **Equation 4.2** are plotted against the calculated binding energies for all eight ligands, and we observe an 89.4% correlation. The excellent agreement between our predicted energies and the experimental binding affinity shown in **Figure 4.5** provides confidence that we predicted a very reasonable binding site for rimonabant.

Discovery of novel CB1-targeting inverse agonists

With a reliable rimonabant binding site pharmacophore for CB1, which is supported by site-directed mutagensis and SAR data, we developed a new, more selective CB1 inverse agonist. After an informed search in PubChem (53), we selected Zinc08587042, which we will refer to as **MSC1** (**Figure 4.6**), a ligand with a 68% similarity to rimonabant according to the 2D Tanimoto coefficient (54). Rather than relying on PubChem's search algorithm, which looks for molecules based on a 0.90 Tanimoto similarity score with respect to rimonabant, we used the predicted CB1 binding site and searched for a ligand that maintained key structural aspects of rimonabant and additionally took advantage of underused residues in the binding site. For example, it was important to find a ligand that had the same amide and pyrazole functional groups to maintain the hydrogen bonds with K3.28 and W5.43 respectively. Yet, we wanted to replace rimonabant's piperidine ring with a phenyl and a substituted polar group to improve the interactions with the aromatic residues and gain an additional hydrogen bond with K7.32, which is in rimonabant's binding site. Furthermore, we wanted to replace the methyl group of the pyrazole group with a polar substitute to form a hydrogen bond with S7.39. **MSC1** is attractive because it contains an acetylphenyl group, which we predicted would reach into the aromatic pocket of the extracellular end, and a triazole, which would replace a methyl group with a nitrogen atom. However, we also wanted a small molecule that was commercially available and that had not been previously tested with CB1 in bioactivity assays. **MSC1** met all of the above criteria.

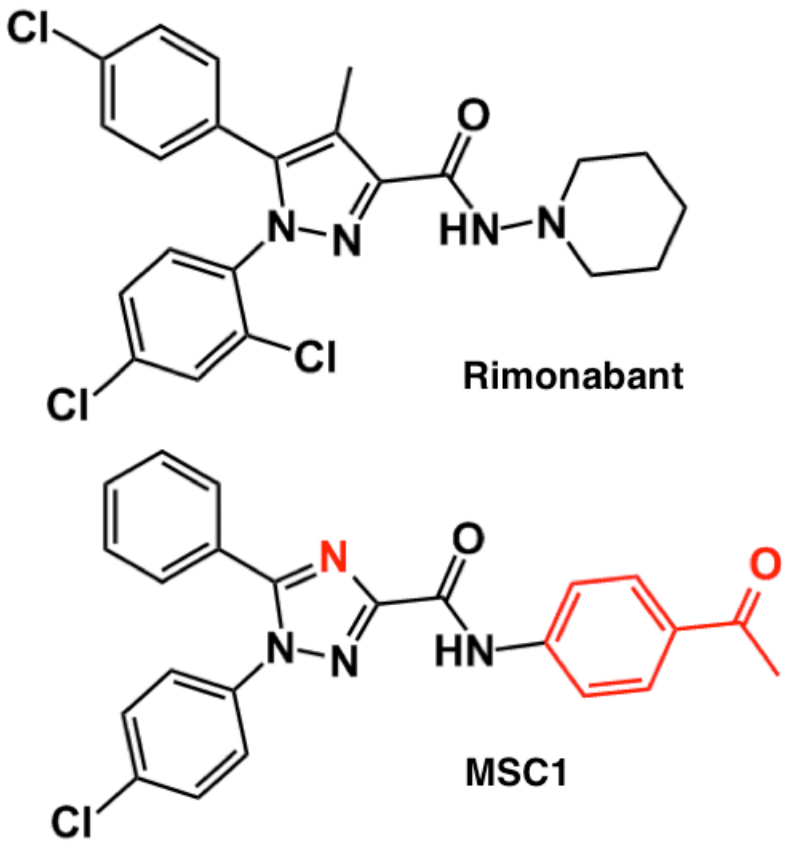


Figure 4.6 Proposed CB1 inverse agonist **MSC1** (Zinc08587042). Structure comparison of rimonabant (top) with **MSC1** (bottom). **MSC1** has a 2D structure that is 68% similar to rimonabant's. Portions of **MSC1** highlighted in red are different from the corresponding parts in rimonabant.

PubChem subsequently identified 15 small molecule ligands that have a 2D structural similarity Tanimoto score of 0.90, or have 2D structures that are 90% similar to that of **MSC1** (**Figure 4.7**). We plan to dock these 16 ligands to the ensemble of predicted CB1 structures, find their corresponding binding energies, and predict their respective pK_i values using the relationship in **Figure 4.5.** Of the 16 MSC small molecules, 14 of them, **MSC1 - MSC6**, **MSC8 – MSC11**, and **MSC13 – MSC16** are commercially available.

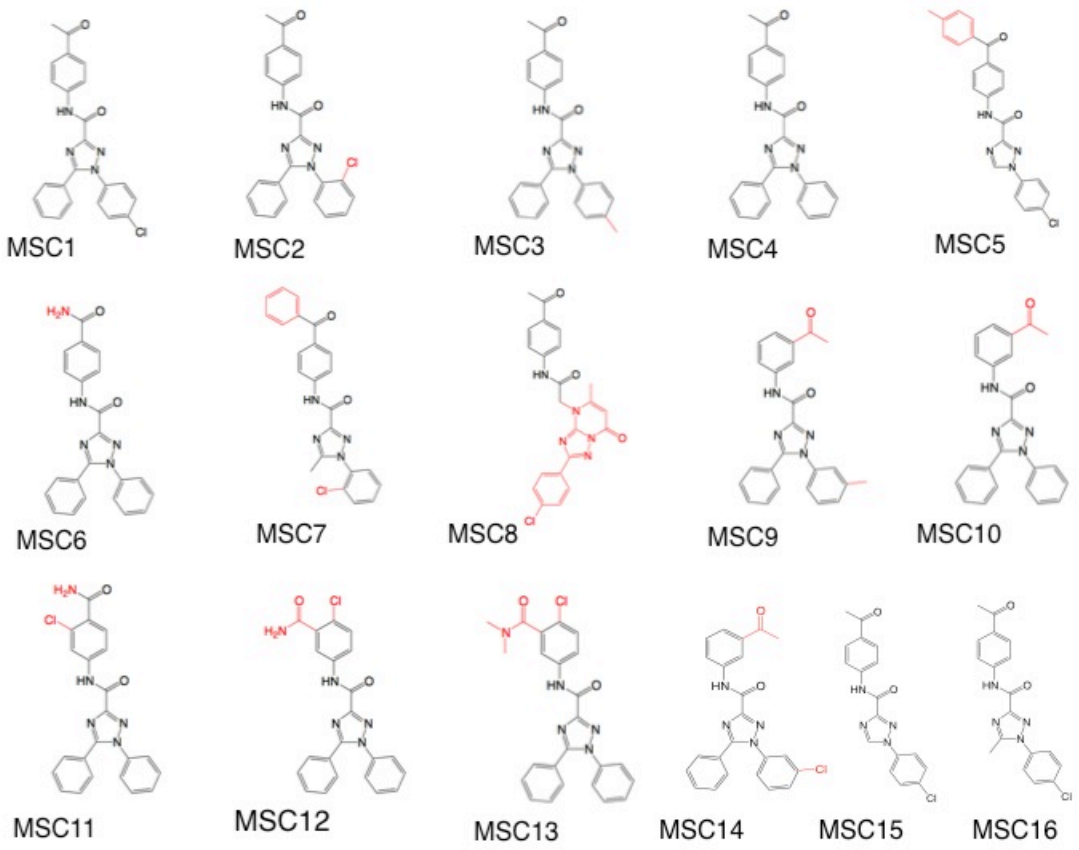


Figure 4.7 Proposed CB1-selective inverse agonists based on PubChem (53) similarity search with MSC1. PubChem identified 15 ligands that have 2D structures similar to MSC1.

CONCLUSIONS

The CB1 receptor inverse agonist rimonabant, along with its structurally related molecules, were docked to previously predicted CB1 receptor conformations. The CB1-rimonabant complex explains site-directed mutagenesis because this inverse agonist is anchored to the receptor via a hydrogen bond with W5.43 rather than just hydrophobic interaction. We confirmed our receptor structure with SAR data, involving rimonabant-related inverse agonists, and our calculated binding energies for the various inverse agonists show excellent, 93%, agreement with the experimentally determined binding affinities.

Understanding how rimonabant interacts with the receptor, we searched for ligands with specific traits that we predicted would enhance its affinity for the binding site. We should note that we did not rely on a PubChem search algorithm based solely on the Tanimoto coefficient to identify potential CB1 inverse agonists. We analyzed rimonabant's binding site in CB1 and predicted

which residues are not being engaged and how structural alterations in rimonabant could maximize interactions with residues in the receptor. We have selected a small molecule that met specific criteria and found another 15 based on a PubChem similarity search. Of these 16 small molecules, 14 ligands are commercially available making it easier for experimentalists to verify our predictions. Our next step is to dock these proposed ligands to calculate the binding energies and provide an estimation of the experimental binding affinities.

It is extremely important to learn if MSC1 and the other small molecules discussed above are CB1-selective strong binding inverse agonists because they could be new candidates for anti-obesity drugs that target CB1. Rimonabant, while effective, is not on the market due to its side effects. Our proposed new inverse agonists based on a realistic binding site might bind almost exclusively to CB1, decreasing the probability of undesirable side effects resulting from off-target interactions.

REFERENCES

1. Crocker E, Eilers M, Ahuja S, Hornak V, Hirshfeld A, Sheves M, Smith SO (2006) Location of Trp265 in metarhodopsin II: Implications for the activation mechanism of the visual receptor rhodopsin. J Mol Biol **357**:163-172.
2. Choe HW, Kim YJ, Park JH, Morizumi T, Pai EF, Krauss N, Hofmann KP, Scheerer P, Ernst OP (2011) Crystal structure of metarhodopsin II. Nature **471**:651-655.
3. Park JH, Scheerer P, Hofmann KP, Choe H-W, Ernst OP (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature **454**:183-187.
4. Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP, Ernst OP (2008) Crystal structure of opsin in its G-protein-interacting conformation. Nature **455**:497-502.
5. Standfuss J, Edwards PC, D'Antona A, Fransen M, Xie G, Oprian DD, Schertler GFX (2011) The structural basis of agonist-induced activation in constitutively active rhodopsin. Nature **471**:656-660.
6. Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK (2011) Crystal structure of the β_2 adrenergic receptor-G_s protein complex. Nature **477**:549-555.
7. Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK (2011) Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. Nature **469**:175-180.
8. Lebon G, Bennett K, Jazayeri A, Tate CG (2011) Thermostabilisation of an agonist-bound conformation of the human adenosine A_{2A} receptor. J Mol Biol **409**:298-310.
9. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG (2011) Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. Nature **474**:521-525.
10. Xu F, Wu HX, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC (2011) Structure of an agonist-bound human A_{2A} adenosine receptor. Science **332**:322-327.

11. Liu W, Chun E, Thompson AA, Chubukov P, Xu F, Katritch V, Han GW, Roth CB, Heitman LH, Ijzerman AP, Cherezov V, Stevens RC (2012) Structural basis for allosteric regulation of GPCRs by sodium ions. Science **337**:232-236.
12. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Huebner H, Pardon E, Valant C, Sexton PM, Christopoulos A, Felder CC, Gmeiner P, Steyaert J, Weis WI, Garcia KC, Wess J, Kobilka BK (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. Nature **504**:101-106.
13. White JF, Noinaj N, Shibata Y, Love J, Kloss B, Xu F, Gvozdenovic-Jeremic J, Shah P, Shiloach J, Tate CG, Grisshammer R (2012) Structure of the agonist-bound neurotensin receptor. Nature **490**:508-513.
14. Zhang J, Zhang K, Gao Z-G, Paoletta S, Zhang D, Han GW, Li T, Ma L, Zhang W, Muller CE, Yang H, Jiang H, Cherezov V, Katritch V, Jacobson KA, Stevens RC, Wu B, Zhao Q (2014) Agonist-bound structure of the human P2Y12 receptor. Nature **509**:119-122.
15. Rinaldi-Carmona M, Barth F, Heaulme M, Shire D, Calandra B, Congy C, Martinez S, Maruani J, Neliat G, Caput D, Ferrara P, Soubrie P, Breliere JC, Lefur G (1994) SR141716A, a potent and selective antagonist of the brain cannabinoid receptor. FEBS Lett **350**:240-244.
16. Scott CE, Abrol R, Ahn KH, Kendall DA, Goddard WA, III (2013) Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations. Protein Sci **22**:101-113.
17. Ahn KH, Scott CE, Abrol R, Goddard WA, III, Kendall DA (2013) Computationally-predicted CB1 cannabinoid receptor mutants show distinct patterns of salt-bridges that correlate with their level of constitutive activity reflected in G protein coupling levels, thermal stability, and ligand binding. Proteins: Struct, Funct, Bioinf **81**:1304-1317.
18. Deupi X, Kobilka BK (2010) Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. Physiology **25**:293-303.
19. Kenakin T, Miller LJ (2010) Seven transmembrane receptors as shapeshifting proteins: The impact of allosteric modulation and functional selectivity on new drug discovery. Pharmacol Rev **62**:265-304.
20. Kobilka BK, Deupi X (2007) Conformational complexity of G-protein-coupled receptors. Trends Pharmacol Sci **28**:397-406.
21. McAllister SD, Rizvi G, Anavi-Goffer S, Hurst DP, Barnett-Norris J, Lynch DL, Reggio PH, Abood ME (2003) An aromatic microdomain at the cannabinoid CB_1 receptor constitutes an agonist/inverse agonist binding region. J Med Chem **46**:5139-5152.
22. Sitkoff DF, Lee N, Ellsworth BA, Huang Q, Kang L, Baska R, Huang Y, Sun C, Pendri A, Malley MF, Scaringe RP, Gougoutas JZ, Reggio PH, Ewing WR, Pelleymounter MA, Carlson KE (2011) Cannabinoid CB1 receptor ligand binding and function examined through mutagenesis studies of F200 and S383. Eur J Pharmacol **651**:9-17.
23. Hurst DP, Lynch DL, Barnett-Norris J, Hyatt SM, Seltzman HH, Zhong M, Song ZH, Nie JJ, Lewis D, Reggio PH (2002) N-(Piperidin-1-yl)-5-(4-chlorophenyl)-1-(2,4-dichlorophenyl)-4-methyl-1H -pyrazole-3-carboxamide (SR141716A) interaction with LYS 3.28(192) is crucial for its inverse agonism at the cannabinoid CB1 receptor. Mol Pharmacol **62**:1274-1287.
24. Chin CN, Lucas-Lenard J, Abadji V, Kendall DA (1998) Ligand binding and modulation of cyclic AMP levels depend on the chemical nature of residue 192 of the human cannabinoid receptor 1. J Neurochem **70**:366-373.
25. Song ZH, Bonner TI (1996) A lysine residue of the cannabinoid receptor is critical for receptor recognition by several agonists but not WIN55212-2. Mol Pharmacol **49**:891-896.
26. Picone RP, Khanolkar AD, Xu W, Ayotte LA, Thakur GA, Hurst DP, Abood ME, Reggio PH, Fournier DJ, Makriyannis A (2005) (-)-7 '-isothiocyanato-11-hydroxy-1 ',1-'

dimethylheptylhexahydrocannabinol (AM841), a high-affinity electrophilic ligand, interacts covalently with a cysteine in helix six and activates the CB1 cannabinoid receptor. Mol Pharmacol **68**:1623-1635.

27. Kapur A, Hurst DP, Fleischer D, Whitnell R, Thakur GA, Makriyannis A, Reggio PH, Abood ME (2007) Mutation studies of Ser7.39 and Ser2.60 in the human CB(1) cannabinoid receptor: Evidence for a serine-induced bend in CB(1) transmembrane helix 7. Mol Pharmacol **71**:1512-1524.
28. Kapur A, Samaniego P, Thakur GA, Makriyannis A, Abood ME (2008) Mapping the structural requirements in the CB1 cannabinoid receptor transmembrane helix II for signal transduction. J Pharmacol Exp Ther **325**:341-348.
29. Fay JF, Dunham TD, Farrens DL (2005) Cysteine residues in the human cannabinoid receptor: Only C257 and C264 are required for a functional receptor, and steric bulk at C386 impairs antagonist SR141716A binding. Biochemistry **44**:8757-8769.
30. Shim J-Y, Bertalovitz AC, Kendall DA (2012) Probing the interaction of SR141716A with the CB1 receptor. J Biol Chem **287**:38741-38754.
31. Hurst D, Umejiego U, Lynch D, Seltzman H, Hyatt S, Roche M, McAllister S, Fleischer D, Kapur A, Abood M, Shi S, Jones J, Lewis D, Reggio P (2006) Biarylpyrazole inverse agonists at the cannabinoid CB1 receptor: Importance of the C-3 carboxamide oxygen/lysine3.28(192) interaction. J Med Chem **49**:5969-5987.
32. D'Antona A, Ahn KH, Kendall DA (2006) Mutations of CB_1 T210 produce active and inactive receptor forms: Correlations with ligand affinity, receptor stability, and cellular localization. Biochemistry **45**:5606-5617.
33. Floriano WB, Vaidehi N, Zamanakos G, Goddard WA, III (2004) HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. J Med Chem **47**:56-71.
34. Goddard WA, III, Kim S-K, Li Y, Trzaskowski B, Griffith AR, Abrol R (2010) Predicted 3D structures for adenosine receptors bound to ligands: Comparison to the crystal structure. J Struct Biol **170**:10-20.
35. Kim S-K, Li Y, Abrol R, Heo J, Goddard WA, III (2011) Predicted structures and dynamics for agonists and antagonists bound to serotonin 5-HT2B and 5-HT2C receptors. J Chem Inf Model **51**:420-433.
36. Kim S-K, Riley L, Abrol R, Jacobson KA, Goddard WA, III (2011) Predicted structures of agonist and antagonist bound complexes of adenosine A(3) receptor. Proteins: Struct, Funct, Bioinf **79**:1878-1897.
37. Nair N, Kudo W, Smith MA, Abrol R, Goddard WA, III, Reddy VP (2011) Novel purine-based fluoroaryl-1,2,3-triazoles as neuroprotecting agents: Synthesis, neuronal cell culture investigations, and CDK5 docking studies. Bioorg Med Chem Lett **21**:5649-5649.
38. Lange JHM, Kruse CG (2005) Medicinal chemistry strategies to CB1 cannabinoid receptor antagonists. Drug Discovery Today **10**:693-702.
39. Salo OMH, Lahtela-Kakkonen M, Gynther J, Jarvinen T, Poso A (2004) Development of a 3D model for the human cannabinoid CB1 receptor. J Med Chem **47**:3048-3057.
40. Silvestri R, Cascio MG, La Regina G, Piscitelli F, Lavecchia A, Brizzi A, Pasquini S, Botta M, Novellino E, Di Marzo V, Corelli F (2008) Synthesis, cannabinoid receptor affinity, and molecular modeling studies of substituted 1-aryl-5-(1H-pyrrol-1-yl)-1H-pyrazole-3-carboxamides. J Med Chem **51**:1560-1576.
41. Abrol R, Griffith AR, Bray JK, Goddard WA, III. Structure prediction of G protein-coupled receptors and their ensemble of functionally important conformations. In: Vaidehi N, Klein-Seetharaman J, Eds. (2012) Membrane Protein Structure: Methods and Protocols. Humana Press, New York, NY.

42. Abrol R, Bray JK, Goddard WA, III (2012) BiHelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. Proteins: Struct, Funct, Bioinf **80**:505-518.
43. Bray JK, Goddard WA, III (2008) The structure of human serotonin 2c G-protein-coupled receptor bound to agonists and antagonists. J Mol Graphics Modell **27**:66-81.
44. Bray JK, Abrol R, Goddard WA, III, Trzaskowski B, Scott CE (2014) SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations. Proc Natl Acad Sci USA **111**:E72-E78.
45. Kam VWT, Goddard WA, III (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. J Chem Theory Comput **4**:2160-2169.
46. Maestro, version 7.5. (2006). Schrodinger, LLC, New York, NY.
47. MacroModel, 9.5. (2007). Schrodinger, LLC, New York, NY.
48. Alkorta I, Alvarado M, Elguero J, Garcia-Granda S, Goya P, Luisa Jimeno M, Menendez-Taboada L (2009) The structure of Rimonabant in the solid state and in solution: An experimental and theoretical study. Eur J Med Chem **44**:1864-1869.
49. Ghosh A, Rapp CS, Friesner RA (1998) Generalized born model based on a surface integral formulation. J Phys Chem B **102**:10983-10990.
50. Lim KT, Brunett S, Iotov M, McClurg RB, Vaidehi N, Dasgupta S, Taylor S, Goddard WA (1997) Molecular dynamics for very large systems on massively parallel computers: The MPSim program. J Comput Chem **18**:501-521.
51. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol **161**:269-288.
52. Kim R, Skolnick J (2008) Assessment of programs for ligand binding affinity prediction. J Comput Chem **29**:1316-1331.
53. Bolton E, Wang Y, Thiessen P, Bryant S. PubChem: Integrated platform of small molecules and biological activities. (2008) Annual Reports in Computational Chemistry. Elsevier Ltd., Oxford, UK, pp. 217-240.
54. Chen R, Weng ZP (2002) Docking unbound proteins using shape complementarity, desolvation, and electrostatics. Proteins: Struct, Funct, Genet **47**:281-294.
55. Ballesteros JA, Weinstein H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In: Sealfon SC, Ed. (1995) Receptor Molecular Biology. Academic Press, Waltham, Massachusetts, pp. 366-428.
56. Scrima M, Di Marino S, Grimaldi M, Mastrogiacomo A, Novellino E, Bifulco M, D'Ursi AM (2010) Binding of the hemopressin peptide to the cannabinoid CB1 receptor: Structural insights. Biochemistry **49**:10449-10457.
57. Audet M, Bouvier M (2012) Restructuring G-protein-coupled receptor activation. Cell **151**:14-23.
58. Bouvier M (2013) Unraveling the structural basis of GPCR activation and inactivation. Nat Struct Mol Biol **20**:539-541.
59. D'Antona A, Ahn KH, Wang L, Mierke DF, Lucas-Lenard J, Kendall DA (2006) A cannabinoid receptor 1 mutation proximal to the DRY motif results in constitutive activity and reveals intramolecular interactions involved in receptor activation. Brain Res **1108**:1-11.

Chapter V

CB1 RECEPTOR STRUCTURAL CHANGES UPON DYNAMICS WITH AND WITHOUT THE PRESENCE OF AN INVERSE AGONIST

Adapted partially from Anton Proposal entitled "Activation of the G_i and G_s proteins via coupling with the human cannabinoid type 1 (CB1) G protein-coupled receptor;" submitted in 2013.

ABSTRACT

G protein-coupled receptors (GPCRs), like the human cannabinoid type 1 (CB1) receptor, are flexible macromolecules that can bind to various intracellular receptors and ligands, but little is known about the ensemble of possible conformations due to the scant number of crystallized GPCRs. Here, we used NAnoscale Molecular Dynamics (NAMD) version 2.9 software to run 50 ns NPT molecular dynamics (MD) simulations on the predicted CB1 receptor in the apo and inverse agonist rimonabant-bound forms inserted into solvated lipid bilayers. We find that the apo-CB1 receptor (without the ligand present) is stable with its R3.50+D6.30 salt bridge intact for 50 ns. We also see the formation of a salt bridge between R148 of the first intracellular loop and D6.25. We expect that both of these salt bridges would stabilize the inactive form since they would prevent the binding of the G protein to the intracellular end. Surprisingly, the inverse agonist rimonabant bound to a conformation of the CB1 receptor that is missing the R3.50+D6.30 salt bridge. However, this polar contact forms during MD and is stable for the final 40 ns of the simulation. We hypothesize that rimonabant conforms the CB1 receptor to take the inactive conformation in an 'induced fit' manner. However, it is necessary to extend the lengths of these simulations to see if the rimonabant-bound receptor conformation becomes fully inactive and stable.

INTRODUCTION

G protein-coupled receptors (GPCRs) are transmembrane (TM) proteins that mediate intracellular signals. Due to their penchant for existing in the hydrophobic lipid bilayer, they are difficult to isolate and crystallize, but thanks to recent technical and methodological advances, there has been an explosion in the number of crystallized GPCRs. However, only four receptors have been crystallized in the activated form (1-11), and even then, crystallizations of these receptors only capture a snapshot of a very dynamic and flexible macromolecule. GPCRs have the ability to exist in multiple active and inactive conformations (12-15), so molecular dynamics (MD) simulations of these receptors are necessary in helping us to understand their activation mechanism and which different conformations are energetically accessible. **Figure 5.1** shows a schematic illustration inspired by Kobilka and Deupi (14) relating the activity levels of receptor conformations and their respective energies. For the apo-receptor, or receptor without a bound ligand, inactive conformations are the lowest or most favorable in energy and are thereby the most populated according to the Boltzmann distribution. As the receptor activates, it undergoes a

conformational transition indicated by the energetic barrier followed by adoption of the activated conformation, which is higher in energy and therefore less populated and less likely to be obtained with crystallization procedures.

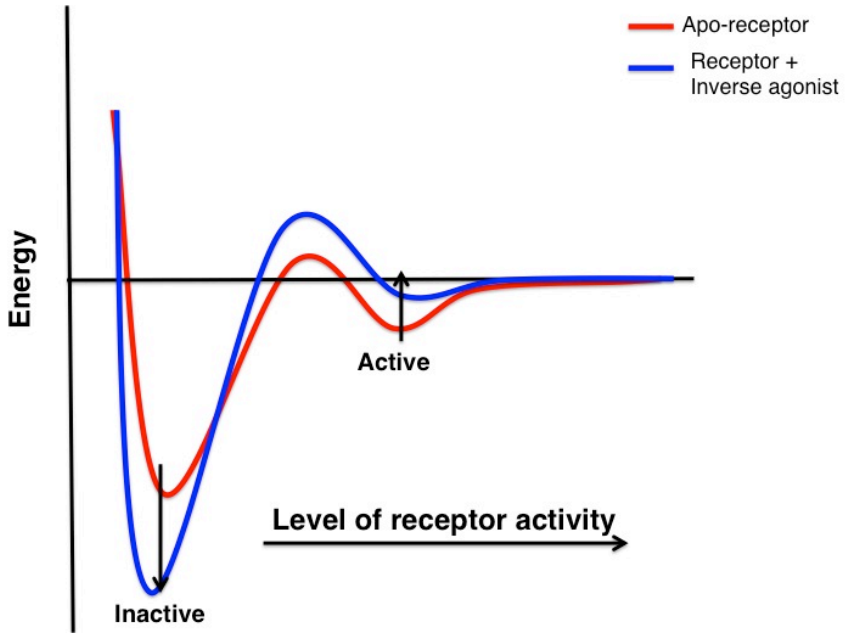


Figure 5.1 Schematic illustration relating energy to receptor level of activation. The red line corresponds to the apo-receptor (receptor alone), and the blue line corresponds to the receptor bound to an inverse agonist. Adapted from **Figure 1** in (14).

When an inverse agonist is added to the receptor, the inactive receptor conformation becomes increasing stabilized while the transition barrier becomes larger, indicating that activation is more energetically difficult. In addition, the activated state becomes higher in energy, so it is less accessible at room temperature. Fluorescence lifetime spectroscopy experiments of the human β_2 adrenergic receptor ($H\beta_2AR$) support this model because the apo-receptor has one conformational state, but in the presence of a neutral antagonist, there are fewer structural fluctuations meaning that this conformation is stabilized (16).

Our aim is to use MD simulations to provide direct information on the atomistic structural changes of the GPCR over time and how these structural changes are impacted by the presence of an inverse agonist. One of the best systems for studying multiple GPCR conformations is the human cannabinoid type 1 (CB1) receptor because it has been experimentally proven to bind to different types of G proteins (17-20) suggesting a structural and functional flexibility, and it has multiple predicted inactive and active conformations (21; 22). It is activated by the psychoactive component of marijuana, and plays an important role in pain and hunger regulation. It is also a

potential target for anti-obesity drugs. But in order to design drugs to interact specifically with CB1, we must understand the receptor structure and the corresponding conformational changes underlying receptor activation that lead to downstream physiological effects.

Our lab routinely carries out MD simulations on large proteins solvated in water with appropriate levels of salt, often on GPCRs (23-39), but also on serine proteases (40), DNA (41-44), RNA (45), peptides on silica surface (46), outer membrane protein A (OmpA) (47), and antifreeze proteins (48). Typically such calculations have ~50,000 atoms per periodic cell. We use such MD software packages include NAMD (49-51), MPSim (52), LAMMPS (53), and AMBER7 (54; 55).

Here, we used NPT MD simulations with NAMD software to understand how our predicted apo-CB1 receptor structure changes in the solvated bilayer, a biologically relevant condition, and in the presence of an inverse agonist. For the apo-receptor, we observe that the structure is stable for 50 ns, but this simulation needs to be lengthened to understand if it adopts a fully inactive conformation. Surprisingly, the inverse agonist rimonabant bound to what appears to be a pre-activated CB1 receptor conformation. However, with dynamics, the receptor starts to form contacts that would resemble an inactive structure. These results suggest that these receptors preferentially adopt the inactive structure, but longer simulations are necessary to confirm this hypothesis.

METHODS

In previous work, we predicted the structures of an ensemble of low energy structures using the GEnSeMBLE method (21; 22) followed by re-docking an optimized version of known CB1-selective inverse agonist rimonabant to the sixth lowest energy structure to obtain a complex with favorable binding energy with strain and ligand solvation (see *Chapter IV*). The apo-receptor structure discussed throughout this chapter is the second lowest CB1 receptor in terms of average energy because this is the lowest energy structure of the ten that contains the R3.50+D6.30 ionic lock. (By convention, salt bridges are indicated by a '+', hydrogen bonds are indicated by a '-', and residues will be referred to with Ballesteros-Weinstein numbering (56).) Here, we add loops and helix 8 before running MD on the apo-receptor or receptor-ligand complex in a solvated and neutralized lipid bilayer.

Adding loops to the TM bundle

The steps below use the all-atom Dreiding force field (FF) (57) unless otherwise noted.

Constructing the loop scaffolds

Loops were added and optimized for the two different receptor structures—one of which contained a bound rimonabant ligand. Each of the six loops were grown in the TM bundle without the other five loops present using a software program called 'growloop' written by Vaclav Cviek of the Goddard lab. The loop scaffolds are initially constructed entirely out of glycine residues with secondary structure dihedral angles for the backbone if indicated (**Table 5.1**). The glycines are grown in one-residue increments from both ends of the start/stop loop residues at the TM termini. We used a tolerance distance of 0.4 Å per residue with a bump cut-off distance of 2.5 Å. The tolerance is the amount of 'space' allotted per residue, so that the distance between the two growing ends of the loop does not exceed *m* (tolerance) with *m* = number of residues left to grow in the loop. The bump cut-off is the acceptable distance between the loop and TM backbone atoms. The number of conformations that were generated for each loop was $10n^2$, with *n* = number of residues in the loop. Secondary structure of the loops was predicted with online servers: Porter Protein Secondary Structure Prediction Server (58; 59), APSSP2: Advanced Protein Secondary Structure Prediction Server (60; 61), and PSIPRED: Protein Prediction Server (62-64). The lengths of TM2 and TM3 were shortened by two residues each on the extracellular (EC) ends in order to generate enough EC1 loops that could span the distance between the TMs without being too strained.

Table 5.1 CB1 loop building. The respective start and stop residues and respective loop sequences for each of the three intracellular (IC) and three extracellular (EC) loops are shown. Secondary structure prediction (SSP) for the loop backbone based on three online servers (58-64) is shown with S=beta sheet and C=random coil. [a]=Modified start residue for EC1 loop. [b]= Modified stop residue for EC1 loop. Originally they were 183 and 186 respectively, but the program could not generate enough loops because of too much strain. *= 'dummy' residue, place holder for the seven-residue EC2-bulge sequence.

Loop	Start Residue	Stop Residue	Sequence	SSP
IC1	144	149	SRSLRC	None
EC1	181[a]	188[b]	HRKDSRNV	None
IC2	217	229	SIHRPLAYKRIVT	None
EC2	255	271	WNCA*CSDIFPHI	None
EC2-bulge	258	263	EKLQSV	None
IC3	309	322	IQRGTQKSIIIHTSEDGKVQVTRP	CCCCCSSSSSSSSCCCSSSSSSCC
EC3	369	375	GKMNKLI	None

Modifying the loop scaffolds

In a subsequent script, written by Caitlin E. Scott and Professor Ravinder Abrol, the individual loops are merged with the TM bundle and modified to have the correct target sequence. First, the individual loop plus the four terminal residues at the loop-TM junction in the TM bundle are minimized 10 steps with SGB solvation (65) using MPSim software (52), and the top 10%, according to energy, are kept for subsequent steps. Then, the glycine, proline, and cysteine residues, which affect the shape of the loop backbone (66), are mutated into the loops with SCREAM software (67). Alanines replace all the other glycine residues. The loop regions and the terminal TM residues are minimized for 10 steps with SGB solvation. The correct sequence is mutated with SCREAM software into the loop scaffolds before the loops and adjacent TM residues are minimized again for 10 steps with SGB solvation. The top 10% of conformations are chosen according to energy, and these loops and the adjoining terminal TM residues are minimized 50 steps with SGB solvation. Any loop-receptor structure that is greater than 50 kcal/mol from the lowest energy structure is discarded from subsequent steps listed below.

Building EC2 loop and an intra-loop disulfide bridge

The EC2 loop was particularly challenging to construct given the presence of a disulfide bridge (DSB) between cysteine residues 257 and 264 (68). To ensure that the two cysteine residues are close enough to form an intra-loop DSB, we replace the seven residues separating the cysteines with a 'dummy' alanine residue. The procedure for inserting the EC2 loop scaffold into the TM bundle follows the same method outlined above, with some exceptions. After attaching the loop scaffold to the TM bundle, the distance between the alpha-carbons of the DSB cysteine residues is calculated. If the distance is greater than 7.5 Å, the loop is discarded. Second, after keeping the top 10% of loops according to energy and mutating the cysteine, proline, glycine, and alanine residues with SCREAM, the distance between the two sulfur atoms in the DSB cysteine residues is calculated. If this distance is greater than 7.0 Å, the loop is discarded. Then, the dummy alanine residue is removed, and the DSB is formed. The charges are corrected for the oxidized cysteines in the DSB, and these residues and their adjacent neighbors are minimized 10 steps with SGB solvation. Subsequently, the entire loop and the terminal TM residues are minimized 10 steps with SGB solvation before completing the rest of the procedure described above by using SCREAM to replace the alanines with the target sequence residues, minimizing the solvated loops and adjoining residues, keeping the best 10% of low energy structures, and finally minimizing the loops and adjoining residues for 50 steps with SGB solvation.

After constructing the DSB and removing the dummy residue, it is necessary to complete the loop by building the portion referred to here as the 'EC2-bulge,' or the loop residues separating the two cysteines in the DSB. For each of the receptor-EC2 loop portions generated in the step above, we use growloop to generate the EC2-bulge portion and insert it into the receptor-EC2 loop. The bulge residues and the four residues connecting the DSB cysteine residues to the previously constructed loop are minimized for 10 steps with SGB solvation. For the loops with the top 10% according to energy, these residues are mutated with SCREAM software into alanine, cysteine, proline, or glycine residues and minimized 10 steps using SGB solvation. The correct target sequence replaces the alanine residues using the SCREAM program before being minimized 10 steps with SGB solvation. The bulge and adjoining loop residues are minimized for 50 steps with SGB solvation. All the TM bundles with the complete EC2 are ranked according to their energies, and the top (n^2/10) loops or those within 50 kcal/mol of the lowest energy structure are selected for subsequent steps.

Merging the loops and attaching them to the TM bundle

In this section, we will first test combinations of first the EC loops to see which three loop conformations produce the lowest energy combination. EC1 and EC3 are extracted from the TM bundle-loop conformations and attached to the TM bundle containing EC2. All the residues in the loops and any residue in the TM bundle within 5 Å of the loops' rotamers are sampled with SCREAM software. The single point energies of loop portion and close residues of the TM-EC loop bundle are determined in vacuum and with SGB solvation. These portions are then minimized for 10 steps in vacuum and 10 steps with SGB solvation. All the combinations of EC1, EC2, and EC3 loops are sorted and ranked according to these four different energy criteria. For the apo-CB1 receptor, the second lowest energy structure after minimization with SGB solvation was selected for subsequent steps because its energy was within several hundredths of the lowest energy complex, it had a lower RMS force, and it ranked first in the other three energy categories. For the CB1-rimonabant receptor, the structure with the best minimized energy with SGB solvation was selected since this type of energy is the most accurate representation.

The TM bundle with the three ECs acted as input for attaching the three intracellular (IC) loops. The IC loops are extracted from the TM bundle-IC loop complex, so that they are individually attached to the TM bundle containing the EC loops. The IC loop residues and any residue within 5 Å have their various rotamers sampled with SCREAM. Then, the loops and surrounding

residues' single point energies are calculated in vacuum and with SGB solvation. These loops are minimized for 10 steps in vacuum and with SGB solvation, and are ranked according to these four energies. The lowest energy structure after minimization with SGB solvation is selected for MD.

Adding helix 8 to the TM bundle

For helix 8, the helix that lies perpendicular to the helical bundle with half of itself exposed to the solvent and the other half exposed to the bilayer, we used a homology model form the human adenosine A_{2A} receptor (HAA2AR) bound to an agonist (PDB ID: 3QAK) (8) because its TM7 was oriented closest to TM7 in CB1. We aligned the NPXXY motifs of TM7 for the two receptors, then using Maestro software (69) rotated helix 8 so it does not clash with CB1's TM1, and then used SCREAM software to mutate the residues with those that match the CB1 sequence. Then, the residues consisting of helix 8 (#399 to 413) are attached to the TM-loop bundle and minimized 50 steps in vacuum.

CB1 receptor molecular dynamics

MD simulations were conducted on two systems—one of the apo-CB1 GPCR and one of the CB1 receptor bound to the inverse agonist rimonabant. This receptor was inserted into a periodic POPC lipid bilayer in a TIP3P water box with a tetragonal periodic cell 85Å x 85Å x 107Å. We added 15 chloride ions to neutralize the systems and to match physiological pH. The apo-CB1 receptor system consisted of a total of 64,197 atoms including 4918 atoms for the receptor, 19,381 membrane atoms, 39,883 water atoms, and 15 chloride atoms for neutralization (**Table 5.2**). For the CB1-rimonabant system, there are 63,242 total atoms including 4918 atoms for the receptor, 18,416 membrane atoms, 39,842 water atoms, 15 chloride atoms for neutralization, and 51 atoms for the ligand.

Table 5.2 Size of the CB1 MD systems

	Number of atoms					
MD system	**Receptor**	**Membrane**	**Water**	**Ions**	**Ligand**	**Total**
Apo-CB1	4918	19,381	39,883	15	0	64,197
CB1+Rimonabant	4918	18,416	39,842	15	51	63,242

We performed NPT simulations with the NAnoscale Molecular Dynamics (NAMD) version 2.9 software (49-51) with the CHARMM22 FF (70; 71) for the receptor atoms, the CHARM27 FF (72) for the lipids, the AMBER9 FF (73) for the ligands. For the non-bonded FF parameters, we

set the cut-off distance for electrostatic and van der Waals (vdW) calculations to be 12 Å. The pairlist distance is 13.5 Å. At this distance, the energies between pairs of atoms will not be calculated, but this pair will be included in the list of interacting atoms in case these two atoms move closer than the cut-off distance later in the simulation. The switching function was selected, so smoothing functions, not truncated cut-off functions were applied to the electrostatic and vdW forces. The switching distance, or when the smoothing function takes effect, is at 10 Å. We excluded atoms connected by two or three bonds from non-bonding interactions.

First we minimized the system, while the TMs, loops, and ligand (if present) are fixed, and the water, ion, lipid, and helix 8 atoms are movable. Minimization lasted for 5 ps with a timestep of 1 fs. There were 20 timesteps per cycle, or 20 fs, before atoms were reassigned to the non-bonded pair list, or list of atoms to calculate pairwise non-bonded energies. The system had periodic boundary conditions that are defined by three vectors : <85Å 0 0>, <0 85Å 0>, <0 0 107Å>, to match the size of the cell. Water molecules were wrapped, or translated, to the other side of the cell if they exited the cell's boundaries. Particle Mesh Ewald (PME) was applied to electrostatics (74; 75). The PME grid had 87 points in the *x* and *y* directions and 110 points in the *z*-direction. Full electrostatic evaluations and short-range non-bonding interactions occurred every timestep or every femtosecond.

Second, we equilibrated the system using the same set of movable atoms. Dynamics lasted for 500 ps at a timestep of 1 fs. The FF parameters, the PME, and many of the integrator parameters were the same as for those used during minimization. Differences included: full electrostatic evaluations occurred every 4 fs, and non-bonding interactions were calculated every 2 fs rather than every femtosecond as during minimization. To maintain a constant pressure of 1.01325 bar (1 atm) and temperature at 310 K for our NPT simulation, we used the Nosé-Hoover Langevin piston control and Langevin thermostat with Langevin dynamics (76; 77). To maintain a constant pressure, we used a flexible cell with constant area in the x and y directions, thus allowing fluctuations in the *z*-direction. The Langevin piston oscillation period was 200 fs, the Langevin piston decay or damping timescale was 200 fs, and the Langevin piston temperature or noise temperature was equal to the target temperature (310K). We initialized the temperature to equal 0K, and the system was heated to 310K. To maintain a constant temperature, we used a damping coefficient of 5 ps^{-1}, and Langevin dynamics did not apply to the hydrogen atoms. All the atoms, not just the waters', are wrapped around the cell. In this case, we applied a margin of 5 Å to partition the cell into smaller cubes. Trajectory coordinates, unit cell information, and details

about the extended system were written out and saved every picosecond, and energies were written out every 100 fs.

Third, we performed a 5 ps minimization on the entire system including all receptor atoms. Since the size of the box changed during dynamics, in this step the PME grid had 84 points in the *x* and *y* directions and 96 points in the *z*-direction, which was smaller than what it was during the first minimization. Also, the full electrostatic evaluations occurred every 4 fs. All the atoms, not just the waters', were wrapped around the cell, and the cell basis vectors were not defined. Everything else matched the parameters for the first minimization.

Fourth, we performed a simulation of the entire system with every atom movable. There is a 250 ps equilibration period followed by 50 ns of NPT dynamics with a timestep of 1 fs. The parameters for the portion are the same as for the first equilibration period except the PME grid size was smaller, matching the dimensions used for the second minimization step. Here, the cell basis vectors and margin for periodic boundary conditions were not defined. Also, in this step, the cell was kept at a constant area throughout the simulation. Finally, the trajectory coordinates were written out every picosecond rather than every 100 fs.

RESULTS AND DISCUSSION

Apo-CB1 receptor dynamics: Structure is stable; see inactive traits start to form

To ensure the stability of the predicted CB1 structures, we ran 50 ns of MD simulations of the apo-CB1 receptor, second lowest in energy in the ensemble, in a neutralized solvated bilayer. The average root-mean-square deviation (RMSD) per residue during the simulation is rather high at 4.01 Å, but many of the important polar contacts are stable. The loop regions and TM1 are flexible, but the other TMs are stable as indicated by the RMSD per residue shown in **Figure 5.2** and **Table 5.3**. The RMSD per residue averaged over the seven TMs is 2.98 Å, which is much lower than the corresponding value for the six loop regions—6.11 Å. The R3.50+D6.30 salt bridge is intact during the entire simulation, and after 5 ns, a new salt bridge between R148 of IC1 and D6.25 forms (**Figure 5.3**). This R148+D6.25 salt bridge is similar to the R2.37+D6.30 ionic lock observed in the fully inactive T3.46A and L3.43A/D2.63A mutants (21; 22). After some dynamics, the constitutively active CB1 apo-receptor starts to resemble the fully inactive T3.46A mutant structure since there are two salt bridges that span the interior of the bundle. This

R148+D6.25 salt bridge would also be effective at preventing G protein coupling to the CB1 receptor and subsequent activation.

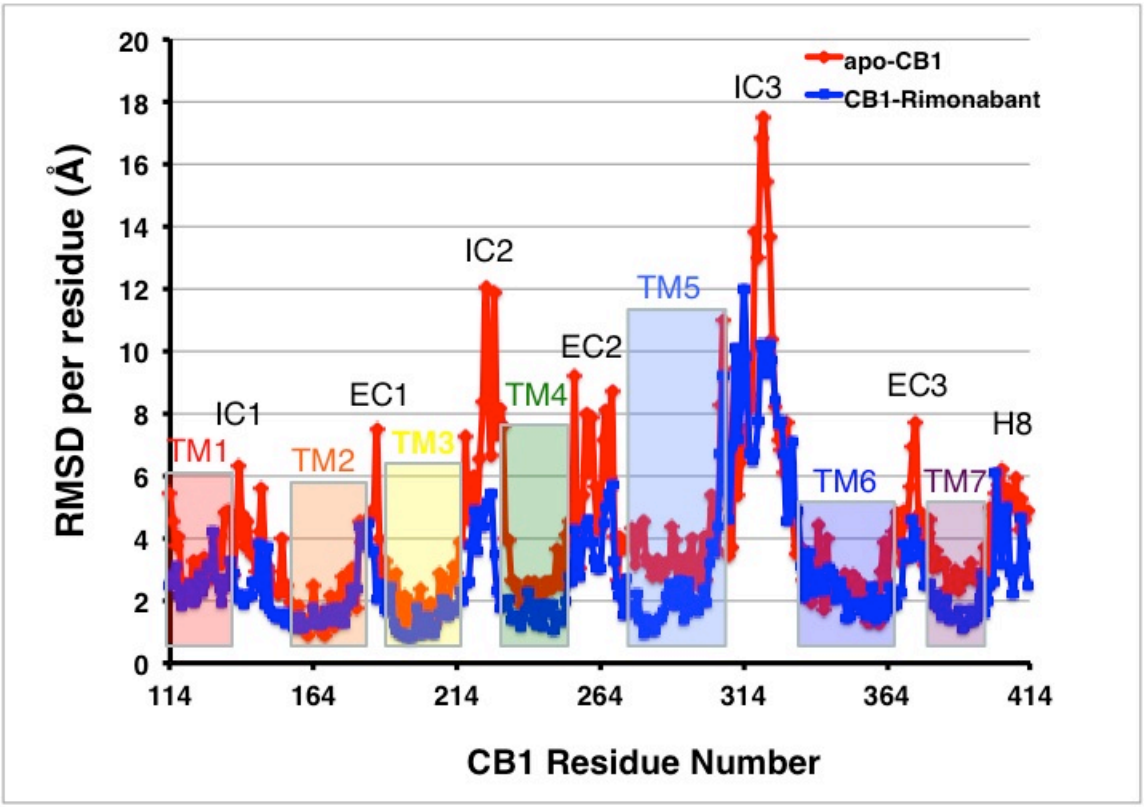


Figure 5.2 RMSD per residue (Å) plotted against all residues in the CB1 receptor. The RMSD values for the apo-CB1 residues are in red, and the corresponding values for the CB1-rimonabant residues are in blue. The seven TMs, three intracellular loops (ICs), three extracellular loops (ECs), and helix 8 are indicated on the plot.

Table 5.3 RMSD per residue averaged over the TM, loop, and helix 8 regions for the apo-CB1 receptor and the CB1-rimonabant complex.

TM#	Apo-CB1 Ave. RMSD (Å)	CB1-Rimonabant Ave. RMSD (Å)	Loop	Apo-CB1 Ave. RMSD (Å)	CB1-Rimonabant Ave. RMSD (Å)
1	3.65	2.51	**IC1**	3.74	2.69
2	2.10	1.74	**EC1**	5.43	3.43
3	2.29	1.48	**IC2**	7.53	3.82
4	3.04	1.69	**EC2**	5.77	3.85
5	3.91	2.45	**IC3**	8.81	7.88
6	2.74	2.29	**EC3**	5.39	3.59
7	3.09	1.88	**Helix8**	5.16	3.62
TM average	2.98	2.01	**Loop average**	6.11	4.21

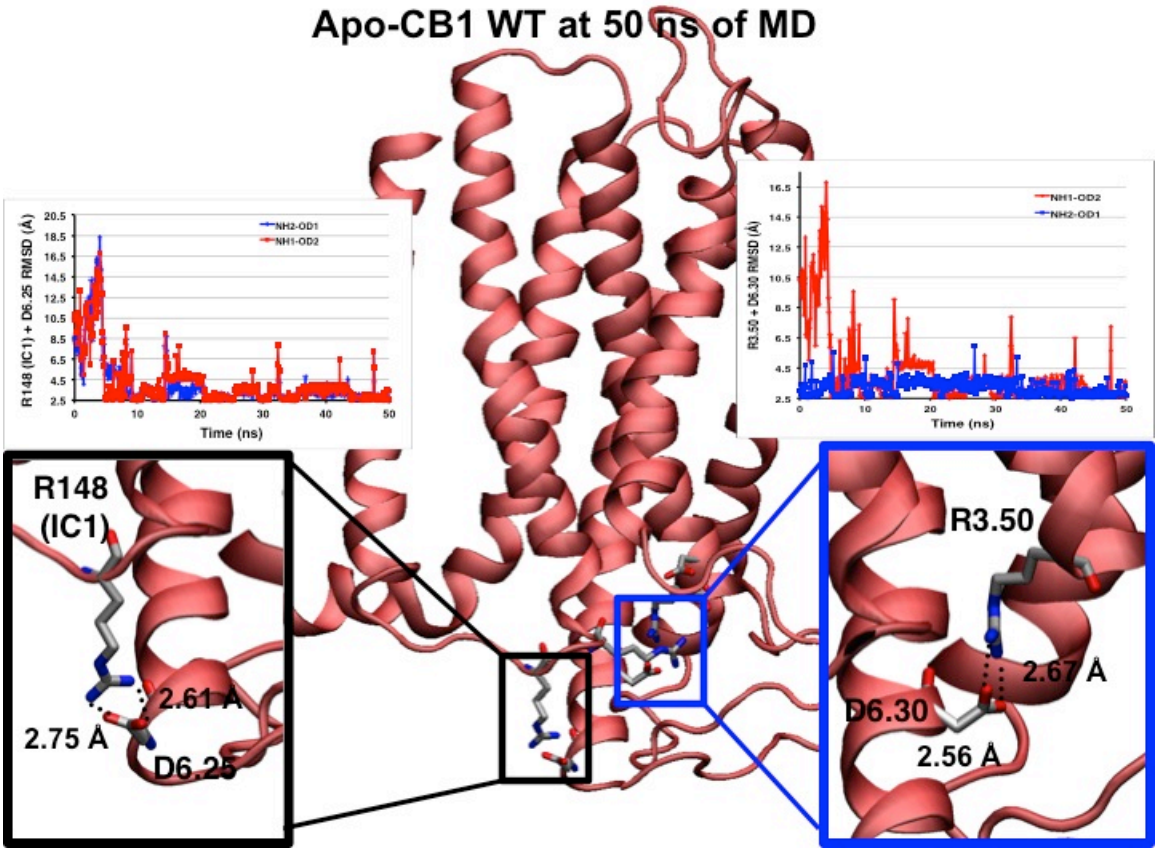


Figure 5.3 Apo-CB1 receptor after 50 ns of MD. Important intracellular salt bridge residues [R148 (IC1), R3.50, D6.25, and D6.30] are labeled in the insets. Salt bridges are indicated with black dotted lines, and their heteroatom distances are labeled. The RMSD distances are plotted with respect to time in the upper graphs for the R148+D6.25 salt bridge (left) and the R3.50+D6.30 salt bridge (right).

Also, the R2.37+D3.49 salt bridge observed in the predicted structures of constitutively active and highly constitutively active receptors is stable for almost 50 ns (**Figure 5.4**). At approximately 49.5 ns, this polar contact breaks. The simulation needs to be extended to see if this salt bridge reforms or if it is permanently broken. Perhaps by breaking this TM2+TM3 salt bridge, there will be an opportunity for the R2.37+D6.30 ionic lock to form. The R148 residue, in close proximity to R2.37, interacts with D6.25 in a stable salt bridge, so the intracellular end of TM2 is being drawn towards TM6 making the formation of the ionic lock highly likely.

This simulation shows that the constitutively active CB1 receptor structure starts to resemble the fully inactive structure, which agrees with previous analysis of Hβ_2AR (12; 14-16), in that in the absence of a stabilizing agonist or G protein, the inactive conformation of the receptor is most energetically favorable and, according to the Boltzmann distribution, most populated.

Surprisingly, we begin to observe such significant conformational changes over a short timescale. It is necessary to run these simulations longer—for a microsecond trajectory perhaps—to see the formation of the R2.37+D6.30 ionic lock and to ensure that these traits are stable.

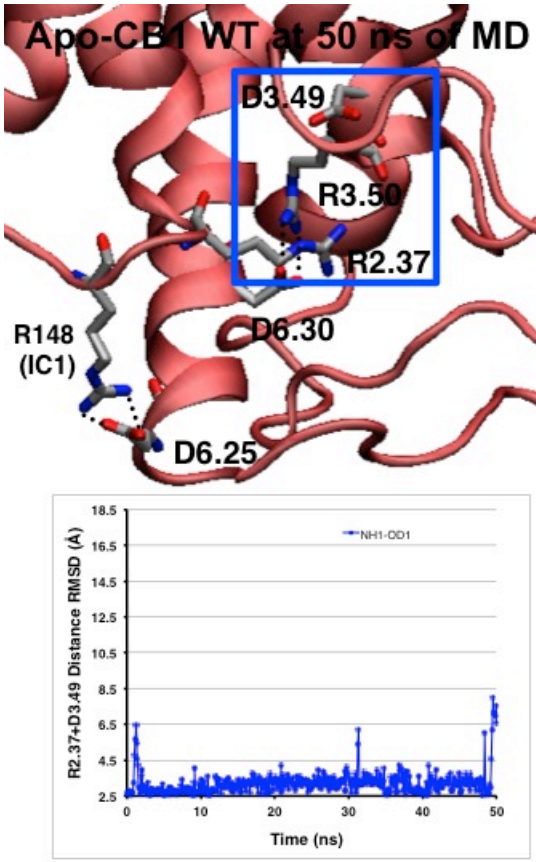


Figure 5.4 Intracellular end of apo-CB1 receptor after 50 ns of MD. Important charged residues [R148 (IC1), R2.37, D3.49, R3.50, D6.25, and D6.30] are labeled. Salt bridges are indicated with black dotted lines. The RMSD distances between the heteroatoms of residues R2.37 and D3.49 are plotted with respect to time in the bottom graph.

CB1-rimonabant dynamics: Inverse agonist induces formation of inactive structure traits

Surprisingly, when we docked the inverse agonist/antagonist rimonabant (78) to the ensemble of low energy constitutively active CB1 structures, the ligand preferentially bound to structures lacking the R3.50+D6.30 ionic lock, which helps stabilize the inactive structure (for more information, see *Chapter IV*). We ran MD on the CB1-rimonabant complex in a physiological system to see how the presence of the inverse agonist affects the constitutively active conformation (**Figure 5.5**). The CB1-rimonabant complex is more stable than the apo-CB1 receptor since its average RMSD per residue is 2.82 Å as compared to 4.01 Å (**Figure 5.2** and **Table 5.3**). The R2.37+D3.49 salt bridge, which broke at the end of the apo-CB1 simulation, is

intact for the 50 ns. After 10 ns, however, a salt bridge between R3.50 and D6.30 forms and is stable for the final 40 ns of the simulation. Much like the apo-CB1 receptor case, certain contacts form, which resemble those indicative of the inactive conformation. Perhaps the presence of the inverse agonist induces the formation of the R3.50+D6.30 ionic lock and other inactive traits. In addition, the formation and breakage of this ionic lock has been observed in long-scale MD simulations of Hβ_2AR. The crystallized form of this receptor bound to the partial inverse agonist carazolol lacks this TM3+TM6 coupling, but this contact forms easily during MD. For the micro-second length simulation of the receptor with and without the inverse agonist bound, the ionic lock forms and breaks multiple times and stays broken for tens of nanoseconds. These data suggest that there are two distinct inactive states for Hβ_2AR—one with the ionic lock and one without (13). We may be observing a similar phenomenon here for the CB1-rimonabant complex in that the receptor can be inactive with or without the intact R3.50+D6.30 ionic lock. The simulations will need to be extended to see if the R2.37+D6.30 ionic lock, which is particularly important for preventing CB1 activation (21; 22), forms and is stable in this complex.

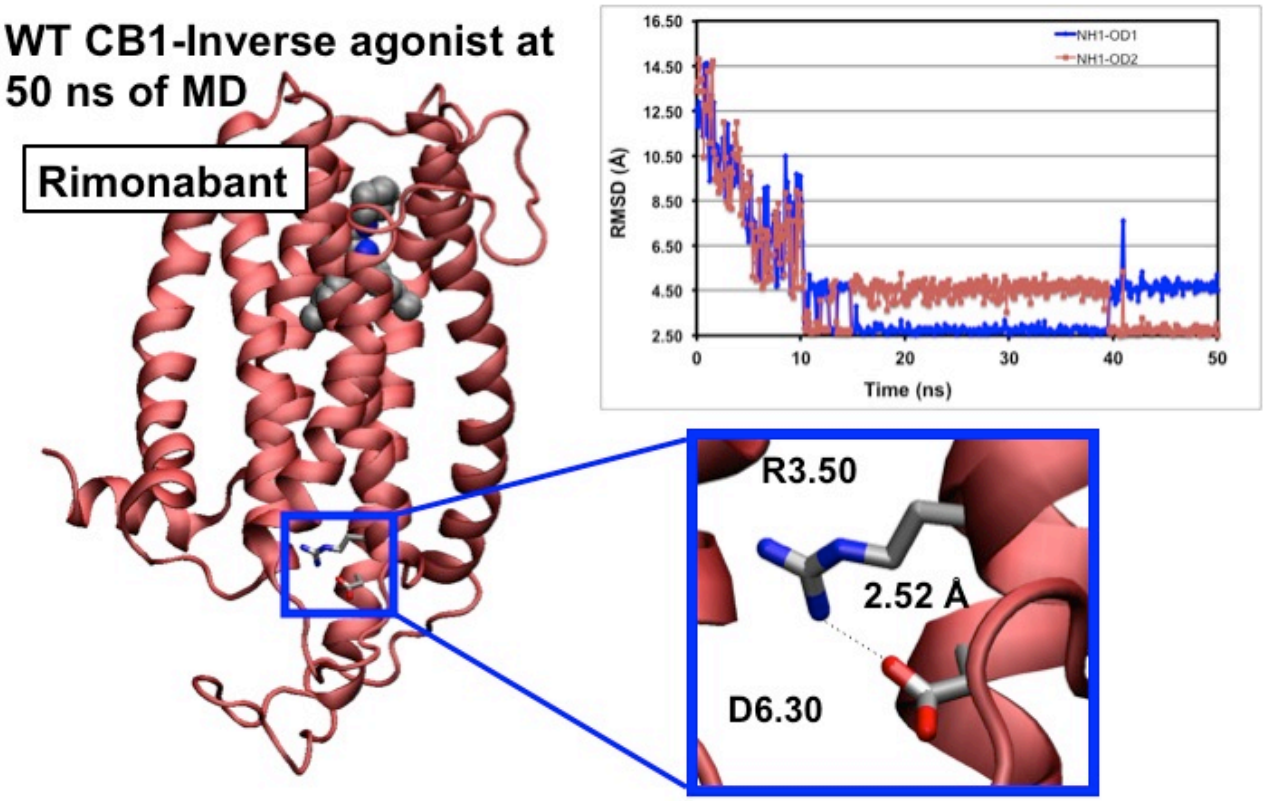


Figure 5.5 CB1-rimonabant complex after 50 ns of MD. Important intracellular salt bridge residues (R3.50and D6.30) are labeled in the inset. The salt bridge is indicated with a black dotted line, and the heteroatom distance is labeled. The RMSD distances are plotted with respect to time in the upper graph for the R3.50+D6.30 salt bridge.

This CB1-rimonabant complex is evidence for both the conformational selection and induced fit theories (discussed in *Chapter I*) (12). We docked ten conformations of the inverse agonist to ten CB1 structures, and the ligand had the lowest binding energy with strain and ligand solvation with the receptor structure that is sixth lowest in energy and missing the R3.50+D6.30 ionic lock. The inverse agonist preferentially binds to a receptor structure that is not only high in energy but looks pre-activated. During the MD simulation, the R3.50+D6.30 ionic lock forms suggesting that rimonabant is inducing the receptor to look inactivated. Longer simulations are needed to see if the receptor becomes fully inactive and if it is stable in that conformation.

Comparison between apo-CB1 and CB1-inverse agonist MD simulations

The CB1 receptor bound to the inverse agonist is more stable during simulations than when it is in its apo form (**Figure 5.2** and **Table 5.3**). Nearly every section of the receptor has a higher RMSD in the apo-receptor than in the inverse agonist-bound form. Surprisingly, the major differences in RMSD are in the loops—especially IC2, which is not near the ligand binding site. The RMSD averaged over all loop residues in IC2 is 7.53 Å in the apo-receptor, but 3.82 Å in the CB1-rimonabant complex. The inverse agonist might have a stabilizing effect on the TM regions and the EC loops, but the connection between rimonabant and its impact of IC2 is not obvious. TM5 of the apo-receptor has a higher RMSD of 3.91 Å compared to 2.45 Å of TM5 of the CB1-rimonabant complex. Rimonabant interacts with TM5, maybe stabilizing it and the connected IC2. The receptor structures themselves, however, are different, and the loops were generated and optimized for the two receptors' structures, so perhaps the cause of the stabilizing effect is from the structure itself, not the inverse agonist. To test this hypothesis, we will need to run MD simulations of the second lowest energy structure (apo form) in complex with rimonabant and of the sixth lowest energy structure (ligand-bound form) without rimonabant present. The results will indicate what effect rimonabant has on these receptors' structures and their relative stabilities. It seems that the CB1 receptor structure of the CB1-rimonabant complex is more stable after MD since the polar contact R3.50+D6.30 is formed, which is absent in the initial conformation. In future work, we will remove rimonabant from the CB1-rimonabant complex and run MD to see if the CB1 structure is stable without the presence of the inverse agonist.

Neutralization issues to address

In this study, we added just the number of chloride ions to neutralize the entire protein, but we have shown in our DNA calculations that it is best to add additional sodium and chloride ions to match the experimental salt concentration (44). For future simulations, we will use TLeap to add

sodium ions to negatively charged residues lacking a salt bridge partner and chloride ions unpartnered with positive residues. Thus, we will neutralize the receptors on a residue-by-residue basis rather than the entire macromolecule.

CONCLUSIONS

We ran MD simulations to see how predicted receptor structures and inverse agonist-bound complexes change over time in biologically relevant conditions. For the apo-CB1 receptor, we see that the intracellular R3.50+D6.30 ionic lock is stable, and another salt bridge between IC1 and D6.25 forms. These observations, along with the breaking of the R2.37+D3.49 salt bridge, might be early signs of the formation of the R2.37+D6.30 ionic lock, which is important for stabilizing the fully inactive form of the receptor. Similarly, the CB1-rimonabant complex forms the TM3+TM6 ionic lock, which is then stable during the simulation. Thus, in the presence of the inverse agonist, the receptor acquires traits that are indicative of the inactive receptor. Perhaps the inverse agonist induces these conformational changes, which would support the conformational selection model (because rimonabant binds preferentially to a pre-activated conformation) and the induced fit model (because after binding, the receptor starts to look inactive). These simulations agree with previous analysis of the Hβ_2AR experiments because we observe two different inactive structures of the receptor bound to an inverse agonist, and the inactive states are more populated even without the presence of an inverse agonist. Longer simulations of the CB1 receptor complexes are necessary to see if they can achieve stable forms of the fully inactive conformations.

REFRENCES

1. Choe HW, Kim YJ, Park JH, Morizumi T, Pai EF, Krauss N, Hofmann KP, Scheerer P, Ernst OP (2011) Crystal structure of metarhodopsin II. Nature **471**:651-655.
2. Park JH, Scheerer P, Hofmann KP, Choe H-W, Ernst OP (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature **454**:183-187.
3. Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP, Ernst OP (2008) Crystal structure of opsin in its G-protein-interacting conformation. Nature **455**:497-502.
4. Standfuss J, Edwards PC, D'Antona A, Fransen M, Xie G, Oprian DD, Schertler GFX (2011) The structural basis of agonist-induced activation in constitutively active rhodopsin. Nature **471**:656-660.
5. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SGF, Choi H-J, DeVree BT, Sunahara RK, Chae PS, Gellman SH, Dror RO, Shaw DE, Weis WI, Caffrey M, Gmeiner P, Kobilka BK (2011) Structure and function of an irreversible agonist-β_2 adrenoceptor complex. Nature **469**:236-240.

6. Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK (2011) Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. Nature **469**:175-180.
7. Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK (2011) Crystal structure of the β_2 adrenergic receptor-G_s protein complex. Nature **477**:549-555.
8. Xu F, Wu HX, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC (2011) Structure of an agonist-bound human A_{2A} adenosine receptor. Science **332**:322-327.
9. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG (2011) Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. Nature **474**:521-525.
10. Lebon G, Bennett K, Jazayeri A, Tate CG (2011) Thermostabilisation of an agonist-bound conformation of the human adenosine A_{2A} receptor. J Mol Biol **409**:298-310.
11. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Huebner H, Pardon E, Valant C, Sexton PM, Christopoulos A, Felder CC, Gmeiner P, Steyaert J, Weis WI, Garcia KC, Wess J, Kobilka BK (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. Nature **504**:101-106.
12. Deupi X, Kobilka BK (2010) Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. Physiology **25**:293-303.
13. Dror RO, Arlow DH, Borhani DW, Jensen MO, Piana S, Shaw DE (2009) Identification of two distinct inactive conformations of the β_2-adrenergic receptor reconciles structural and biochemical observations. Proc Natl Acad Sci USA **106**:4689-4694.
14. Kobilka BK, Deupi X (2007) Conformational complexity of G-protein-coupled receptors. Trends Pharmacol Sci **28**:397-406.
15. Kenakin T (2003) Ligand-selective receptor conformations revisited: The promise and the problem. Trends Pharmacol Sci **24**:346-354.
16. Ghanouni P, Gryczynski Z, Steenhuis JJ, Lee TW, Farrens DL, Lakowicz JR, Kobilka BK (2001) Functionally different agonists induce distinct conformations in the G protein coupling domain of the beta(2) adrenergic receptor. J Biol Chem **276**:24433-24436.
17. D'Antona A, Ahn KH, Wang L, Mierke DF, Lucas-Lenard J, Kendall DA (2006) A cannabinoid receptor 1 mutation proximal to the DRY motif results in constitutive activity and reveals intramolecular interactions involved in receptor activation. Brain Res **1108**:1-11.
18. Glass M, Felder CC (1997) Concurrent stimulation of cannabinoid CB1 and dopamine D2 receptors augments cAMP accumulation in striatal neurons: Evidence for a G(s) linkage to the CB1 receptor. J Neurosci **17**:5327-5333.
19. Abadji V, Lucas-Lenard JM, Chin CN, Kendall DA (1999) Involvement of the carboxyl terminus of the third intracellular loop of the cannabinoid CB1 receptor in constitutive activation of G(s). J Neurochem **72**:2032-2038.
20. Howlett AC, Fleming RM (1984) Cannabinoid inhibition of adenylate-cyclase. Pharmacology of the response in neuro-blastoma cell-membranes. Mol Pharmacol **26**:532-538.
21. Ahn KH, Scott CE, Abrol R, Goddard WA, III, Kendall DA (2013) Computationally-predicted CB1 cannabinoid receptor mutants show distinct patterns of salt-bridges that correlate with their level of constitutive activity reflected in G protein coupling levels, thermal stability, and ligand binding. Proteins: Struct Funct Bioinform **81**:1304-1317.
22. Scott CE, Abrol R, Ahn KH, Kendall DA, Goddard WA, III (2013) Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations. Protein Sci **22**:101-113.

23. Goddard WA, III, Abrol R (2007) Three-dimensional structures of G protein-coupled receptors and binding sites of agonists and antagonists. J Nutr **137**:1528S-1538S.
24. Kirkpatrick A, Heo J, Abrol R, Goddard WA, III (2012) Predicted structure of agonist-bound glucagon-like peptide 1 receptor , a class B G protein-coupled receptor. Proc Natl Acad Sci USA **109**.
25. Kim S-K, Li Y, Abrol R, Heo J, Goddard WA, III (2011) Predicted structures and dynamics for agonists and antagonists bound to serotonin 5-HT2B and 5-HT2C receptors. J Chem Inf Model **51**:420-433.
26. Kim S-K, Li Y, Park C, Abrol R, Goddard WA, III (2010) Prediction of the three-dimensional structure for the rat urotensin II receptor, and comparison of the antagonist binding sites and binding selectivity between human and rat receptors from atomistic simulations. ChemMedChem **5**:1594-1608.
27. Kim S-K, Riley L, Abrol R, Jacobson KA, Goddard WA, III (2011) Predicted structures of agonist and antagonist bound complexes of adenosine A_3 receptor. Proteins **79**:1878-1897.
28. Heo J, Han S-K, Vaidehi N, Wendel J, Kekenes-Huskey PM, Goddard WA, III (2007) Prediction of the 3D structure of FMRF-amide neuropeptides bound to the mouse MrgC11 GPCR and experimental validation. ChemBioChem **8**:1527-1539.
29. Heo J, Ja WW, Benzer S, Goddard WA, III (2008) The predicted binding site and dynamics of peptide inhibitors to the methuselah GPCR from *Drosophila melanogaster*. Biochemistry **47**:12740-12749.
30. Heo J, Vaidehi N, Wendel J, Goddard WAI (2007) Prediction of the 3D structure of rat MrgA G protein-coupled receptor and identification of its binding site. J Mol Graphics Model **26**:800-812.
31. Hummel P, Vaidehi N, Floriano WB, Hall SE, Goddard WA, III (2005) Test of the binding threshold hypothesis for olfactory receptors: Explanation of the differential binding of ketones to the mouse and human orthologs of olfactory receptor 912-93. Protein Sci **14**:703-710.
32. Goddard WA, III, Kim S-K, Li Y, Trzaskowski B, Griffith AR (2010) Predicted 3D structures for adenosine receptors bound to ligands: Comparison to the crystal structure. J Struct Biol **170**:10-20.
33. Jaramillo-Botero A, Nielsen RJ, Abrol R, Su J, Pascal T, Mueller J, Goddard WA, III (2012) First-principles-based multiscale, multiparadigm molecular mechanics and dynamics methods for describing complex chemical processes. Top Curr Chem **307**:1-42.
34. Li Y, Zhu F, Vaidehi N, Goddard WA, III, Sheinerman F, Reiling S, Morize I, Mu L, Harris K, Ardati A, Laoui A (2007) Prediction of the 3D structure and dynamics of human DP G-protein coupled receptor bound to an agonist and an antagonist. J Am Chem Soc **129**:10720-10731.
35. Peng JY-C, Vaidehi N, Hall SE, Goddard WA, III (2006) The predicted 3D structures of the human M1 muscarinic acetylcholine receptor with agonist or antagonist bound. ChemMedChem **1**:878-890.
36. Spijker P, Vaidehi N, Freddolino PL, Hilbers PAJ, Goddard WA, III (2006) Dynamic behavior of fully solvated β2-adrenergic receptor , embedded in the membrane with bound agonist or antagonist. Proc Natl Acad Sci USA **103**.
37. Tan J, Abrol R, Trzaskowski B, Goddard WA, III (2012) 3D structure prediction of TAS2R38 bitter receptors bound to agonists phenylthiocarbamide (PTC) and 6-n-propylthiouracil (PROP). J Chem Inf Model **52**:1875-1885.
38. Trabanino RJ, Vaidehi N, Goddard WA, III (2006) Exploring the molecular mechanism for color distinction in humans. J Phys Chem B **110**:17230-17239.
39. Vaidehi N, Schlyer S, Trabanino RJ, Floriano WB, Abrol R, Sharma S, Kochanny M, Koovakat S, Dunning L, Liang M, Fox JM, Mendonc FLD, Pease JE, Goddard WA, III,

Horuk R (2006) Predictions of CCR1 chemokine receptor structure and BX 471 antagonist binding followed by experimental validation. J Biol Chem **281**:27613-27620.
40. Cho AE, Wendel JA, Vaidehi N, Kekenes-Huskey PM, Floriano WB, Maiti PK, Goddard WA, III (2004) The MPSim-dock hierarchical docking algorithm: Application to the eight trypsin inhibitor cocrystals methods. J Comput Chem **26**:48-71.
41. Maiti PK, Pascal TA, Vaidehi N, Heo J, Goddard WA, III (2006) Atomic-level simulations of seeman DNA nanostructures: The paranemic crossover in salt solution. Biophys J **90**:1463-1479.
42. Maiti PK, Pascal TA, Vaidehi N, Goddard WA, III (2007) Understanding DNA based nanostructures. Journal of Nanoscience and Nanotechnology **7**:1-9.
43. Shin YS, Ahmad H, Shi Q, Kim H, Pascal TA, Fan R, Goddard WAI, Heath JR (2010) Chemistries for patterning robust DNA microbarcodes enable multiplex assays of cytoplasm proteins from single cancer cells. ChemPhysChem **11**:3063-3069.
44. Pascal TA, Goddard WA, III, Maiti PK, Vaidehi N (2012) Role of specific cations and water entropy on the stability of branched DNA motif structures. J Phys Chem B **116**:12159-12167.
45. Zhu R, Janetzko F, Zhang Y, Duin ACTV, Goddard WA, Salahub DR (2008) Characterization of the active site of yeast RNA polymerase II by DFT and ReaxFF calculations. Theor Chem Acc **120**:479-489.
46. McAlpine MC, Agnew HD, Rohde RD, Blanco M, Ahmad H, Stuparu AD, Goddard WA, III, Heath JR (2008) Peptide-nanowire hybrid materials for selective sensing of small molecules. J Am Chem Soc **130**:9583-9589.
47. Pascal TA, Abrol R, Mittal R, Wang Y, Prasadarao NV, Goddard WA, III (2010) Experimental validation of the predicted binding site of *Escherichia coli* K1 outer membrane protein A to human brain microvascular endothelial cells. J Biol Chem **285**:37753-37761.
48. Wang S, Amornwittawat N, Juwita V, Kao Y, Duman JG, Pascal TA, Goddard WA, III, Wen X (2009) Arginine, a key residue for the enhancing ability of an antifreeze protein of the beetle *Dendroides canadensis*. Biochemistry **48**:9696-9703.
49. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem **26**:1781-1802.
50. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graphics **14**:33-38.
51. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) NAMD2: Greater scalability for parallel molecular dynamics. J Comput Phys **151**:283-312.
52. Lim KT, Brunett S, Iotov M, McClurg RB, Vaidehi N, Dasgupta S, Taylor S, Goddard WA, III (1997) Molecular dynamics for very large systems on massively parallel computers: The MPSim program. J Comput Chem **18**:501-521.
53. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys **117**:1-19.
54. Case DA, Pearlman DA, Caldwell JW, Cheatham TE, Wang J, Ross WS, Simmerling C, Darden T, Merz KM, Stanton RV, Cheng A, Vincent JJ, Crowley M, Tsai V, Gohilke H, Radmer R, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC, Weiner P, Kollman PA. 1999. AMBER7 Users' Manual, University of California, San Francisco.
55. Case DA, Cheatham III TE, Darden T, Gohlke H, Luo R, Merz Jr. KM (2005) The Amber biomolecular simulation programs. J Comput Chem **26**:1668-1688.
56. Ballesteros JA, Weinstein H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled

receptors. In: Sealfon SC, Ed. (1995) Receptor Molecular Biology. Academic Press, Waltham, Massachusetts, pp. 366-428.
57. Mayo SL, Olafson BD, Goddard WA (1990) DREIDING: A generic force field for molecular simulations. J Phys Chem **94**:8897-8909.
58. http://distill.ucd.ie/porter/.
59. Pollastri G, McLysaght A (2004) Porter: A new, accurate server for protein secondary structure prediction. Bioinformatics **21**:1719-1720.
60. http://www.imtech.res.in/raghava/apssp2/.
61. Raghava GPS. APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. (2002) CASP5. pp. A-132.
62. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol **292**:195-202.
63. http://bioinf.cs.ucl.ac.uk/psipred/: 2002-2008.
64. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT (2005) Protein structure prediction servers at university college london. Nucleic Acids Res **33**:W36-W38.
65. Ghosh A, Rapp CS, Friesner RA (1998) Generalized born model based on a surface integral formulation. J Phys Chem B **102**:10983-10990.
66. Yohannan S, Faham S, Yang D, Whitelegge JP, Bowie JU (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. Proc Natl Acad Sci USA **101**:959-963.
67. Kam VWT, Goddard WA (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. J Chem Theory Comput **4**:2160-2169.
68. Fay JF, Dunham TD, Farrens DL (2005) Cysteine residues in the human cannabinoid receptor: Only C257 and C264 are required for a functional receptor, and steric bulk at C386 impairs antagonist SR141716A binding. Biochemistry **44**:8757-8769.
69. Maestro, version 7.5. (2006). Schrodinger, LLC, New York, NY.
70. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. J Chem Theory Comput **8**:3257-3273.
71. Mackerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem **25**:1400-1415.
72. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem **4**:187-217.
73. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P (1984) A new force-field for molecular mechanical simulation of nucleic-acids and proteins. J Am Chem Soc **106**:765-784.
74. Darden T, York D, Pedersen L (1993) Particle mesh ewald: An N.log(N) method for ewald sums in large systems. J Chem Phys **98**:10089-10092.
75. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh ewald method. J Chem Phys **103**:8577-8593.
76. Martyna GJ, Tobias DJ, Klein ML (1994) Constant-pressure molecular dynamics algorithms. J Chem Phys **101**:4177-4189.
77. Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) Constant-pressure molecular dynamics simulation: The Langevin piston method. J Chem Phys **103**:4613-4621.
78. Rinaldi-Carmona M, Barth F, Heaulume M, Shire D, Calandra B, Congy C, Martinez S, Maruani J, Neliat G, Caput D, Ferrara P, Soubrie P, Breliere JC, Le Fur G (1994)

SR141716A, a potent and selective antagonist of the brain cannabinoid receptor. FEBS Lett **350**:240-244.

Chapter VI

STRUCTURAL PREDICTION OF THE HUMAN CANNABINOID TYPE 2 (CB2) RECEPTOR AND COMPARISON WITH ITS SISTER CB1 RECEPTOR

The majority of the CB2 WT and T3.46A structure prediction computational experiments were performed by Kyle Tejada as part of the Caltech Freshman Summer Research Institute (FSRI).

The ligand binding assay results shown herein were preformed by Dr. Kwang Hyun Ahn and other members of Professor Debra Kendall's lab.

ABSTRACT

The human cannabinoid type 2 (CB2) receptor is an attractive target for pain medication. Since it is located in the peripheral nervous and immune systems, drugs that act on this receptor are less likely to cause unwanted psychoactive side effects. However, drug design is difficult because CB2 has never been crystallized, and its structure remains unknown. Previously, we predicted the structure of its sister receptor, human cannabinoid type 1 (CB1), using the GEnSeMBLE computational programs and explained the functional differences caused by structural changes resulting from the mutation of a single threonine residue 127 (T3.46) on the intracellular end of the third transmembrane region (TM). In the CB1 receptor, changing this threonine to an alanine (T3.46A) causes the receptor to become fully inactive. Changing this same threonine to an isoleucine (T3.46I) causes the receptor to become highly constitutively active. We predicted, and subsequent experimental evidence supported our claim, that a salt bridge between an arginine on TM2 and an aspartic acid on TM6 stabilizes the inactive CB1 conformation. Using the predicted structure of CB1 WT as a template, we predicted the structures of the CB2 wild-type (WT) receptor plus two threonine mutants (T3.46A, T3.46I). Surprisingly, the resulting CB2 receptor structure predictions are significantly different from its CB1 counterparts. For example, the threonine to alanine (T3.46A) mutant lacked the TM2 and TM6 salt bridge indicating that the CB2 version is constitutively active like the CB2 WT receptor. The threonine to isoleucine (T3.46I) mutant also is missing the TM2 and TM6 salt bridge as well as the signature polar contacts of the highly activated CB1 structures. Thus, we predict all three receptors are constitutively active, which is different from what was predicted for the CB1 receptor and its corresponding mutants. However, subsequent agonist binding assays performed by Professor Debra Kendall and her lab support our conclusions based on our structural predictions. With this information, we propose a double mutant, G2.48V/T3.46A, that we believe will contain the TM2 and TM6 salt bridge and thus will be fully inactive.

INTRODUCTION

There are few options available for the treatment of multiple and chronic types of pain. The human cannabinoind type 1 (CB1) receptor, which is activated by cannabis, is an attractive target for pain medication, but its usefulness is hampered by the psychoactive side effects including hyperthermia, anxiety, sedation, and catalepsy [for review, see (1)]. CB1 is located primarily in the central nervous system (CNS) (2-6) and less significantly in the peripheral nervous system (7-

9). A more attractive target is the human cannabinoid type 2 (CB2) receptor. Unlike its sister receptor, CB2 is expressed to a much smaller extent in the CNS (10) and is located primarily in peripheral nervous system and immune tissues including the spleen and tonsils where its mRNA concentration is 10 to 100 times greater than that of CB1 mRNA (7). Due to limited expression in the CNS, pain medication that targets CB2 will be less likely to have the same psychoactive side effects exhibited by those that interact with CB1. The CB2-selective agonist AM1241 reduces nocieption (sensitization to pain), inflammation, and neuropathic pain in animal models (11-21). The antinocieption effect is reversed by the CB2-selective antagonist AM630, but not by the CB1-selective antagonist AM251 proving that this sensation is mediated by the CB2, not the CB1 receptor (11; 22). Furthermore, agonists like WIN55212-2 and CP55940, which preferentially bind to CB2 but have a significant affinity for CB1, have been proven to cause pain reduction mostly through the CB1 receptor (12; 23-26). Unlike CB2-selective AM1241, WIN55212-2 has also been shown to cause the undesirable CNS side-effects listed above (7), thus indicating the importance for discovering a highly selective-CB2 agonist. As of 2013, a CB2 agonist has not been FDA-approved for pain treatment (27).

While CB2 agonists like AM1241, have shown therapeutic promise, drug design for CB2 has been especially difficult considering that CB2 has never been crystallized, and its structure remains unknown. However, our lab has been developing GEnSeMBLE (GPCR Ensemble of Structures in Membrane BiLayer Environment) methodology to successful predict the structures of GPCRs (28-32) and have applied these techniques to CB2's sister receptor, CB1 (32-34). In previous studies, Professor Debra Kendall and her lab at the University of Connecticut showed that mutating a single threonine residue 210 [T3.46 according to Ballesteros-Weinstein numbering (35)] on the third transmembrane region (TM) causes significant changes in the receptor's function. When this threonine is mutated to an alanine (T3.46A), the receptor shows a significant decrease in binding affinity for the agonist, but a slight increase in binding affinity for the inverse agonist/antagonist. Conversely, when this threonine becomes an isoleucine (T3.46I), the receptor has a slight increase in binding affinity for the agonist, but a significant decrease in binding affinity for inverse agonist/antagonist. Thus, the T3.46A was believed to be the fully inactive form of the receptor, and T3.46I was the constitutively active mutant (36; 37). Subsequent GTPγS assays conducted without the ligands present confirmed these conclusions about the functional states of the receptors (33; 34). Based on our structural predictions of the wild-type (WT) CB1 receptor and these two single mutants, we suggested that a salt bridge between an arginine on TM2, R2.37, and an aspartic acid on TM6, D6.30, stabilizes the inactive

conformation, and subsequent experimental GTPγS assays proved that R2.37 plays an important role in preventing activation (33; 34).

Here, we use a similar procedure to predict the structure of the CB2 receptor as we did for the CB1 receptor. Since T3.46 plays an important role in CB1's structure, it is unique to the cannabinoid family, and its mutation affects CB1's function, we also predicted the structures of CB2 T3.46A and T3.46I. Since there is a large percentage of conserved sequence identity between the two cannabinoid receptors (54.56%, see below), we expected that there would be significant structural similarities and included CB1 along with the experimentally determined crystallized GPCRs to act as starting templates for our CB2 structure prediction. Surprisingly, the predicted CB1 and CB2 WT receptor structures are very different from one another and share very few stabilizing salt bridges or contact patterns. Even more unexpected, mutating the T3.46 residue did significantly impact the structure, but not the function of the receptor as in CB1. All three predicted structures of the CB2 WT and mutants appear to be constitutively active. The R/K2.37 and D6.30 salt bridge that stabilized the fully inactive CB1 T3.46A mutant was not present in the consensus of our ensemble of low energy structures. However, our CB2 predicted structures are supported by agonist binding assays that show the agonist has strong binding affinities for all three receptors. Based on our structural predictions, we propose a double mutant for CB2, G2.48V/T3.46A, that we anticipate will be fully inactive.

METHODS

(For more details about structure prediction methodology, see *Chapter II*.)

We followed a similar approach to predicting the structures of the human CB2 WT receptor and its two mutants as we did with the CB1 receptors (**Figure 6.1**) (28-32). Briefly, we estimated the lengths of the TMs by using PredicTM to identify the seven hydrophobic regions of the amino acid sequence, and by using online servers' secondary structure predictions (SSP) to develop a consensus of the alpha-helical extensions beyond the bilayer. Then there were two different ways to generate the actual TMs—with the program OptHelix or using homology methods with the crystal and CB1 templates. For the OptHelix route, canonical poly-alanine helices were individually constructed to match the lengths of the TMs predicted by PredicTM and SSP. The glycine, proline, serine, and threonine residues were added in with the Side Chain Excitation Analysis Method (SCREAM) software program (38). After a series of molecular dynamics (MD)

and minimization to obtain the optimized helical shape, the remaining residues replace the alanine residues using SCREAM. Then, the seven individual helices are aligned to one of the crystal or predicted CB1 templates. With the homology route, the crystallized or predicted receptor is separated into seven individual helices that are mutated with SCREAM to match the target sequence before being optimized and added back to the receptor bundle. Then both types of helices are used to generate starting structures for the BiHelix program (28), where pairs of the seven helices are rotated in 30º increments. The pairwise interactions between the helices are summed to create an estimate for the energy for the entire bundle. The lowest energy bundles are constructed and minimized in the program ComBiHelix. The lowest total energy structure is starting structure for the SuperBiHelix program (32), which calculates the energies of pairs of helices while sampling local rotation, tilt, and sweep angles (as defined in *Chapter II*). The lowest energy structures estimated from SuperBiHelix are constructed in SuperComBiHelix, and four types of energies are calculated for the entire helical bundle—the total charged energy, the total neutralized energy, the charged interhelical energy, and the neutralized interhelical energy. The structures are ranked according to the average of the four energy ranks, and an ensemble of ten low energy structures are selected for further analysis and ligand docking.

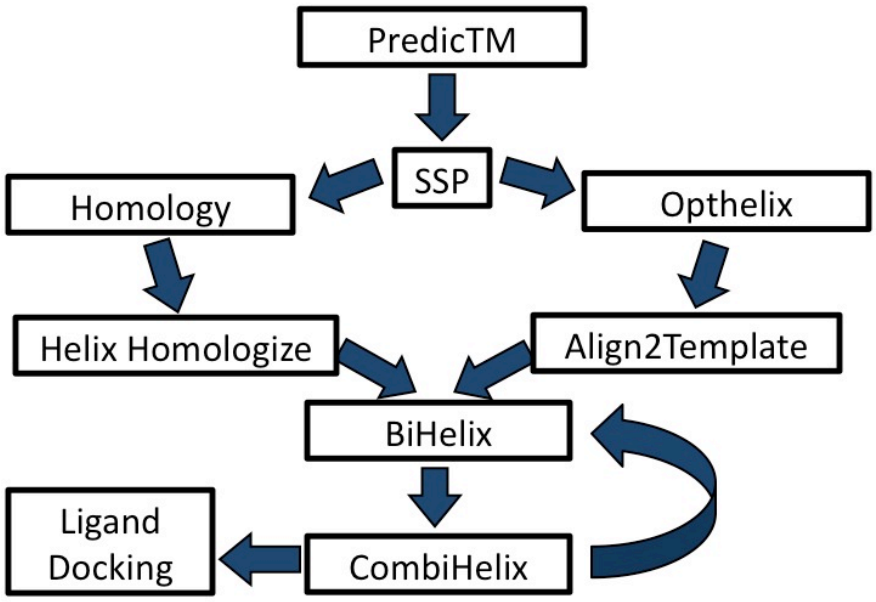


Figure 6.1 Diagram explaining the procedure for predicting the structure of the CB2 receptor and its two mutants. Figure courtesy of Kyle Tejada.

Predicting the CB2 TM lengths

We performed a BLAST (39) search for protein sequences related to CB2. Using the default E-value (0.1) value, we found 1097 sequences ranging in overall sequence identity from 81.11% to 12.22%. The number of sequences found for CB2 is significantly smaller than the number identified for CB1 (1387 sequences ranging in overall sequence identity from 100% to 8.47%). The default E-value was not sufficient for identifying enough related sequences to the target human CB2 sequence, so the BLAST search was repeated with a larger E-value of 1.0, which returned a larger number, 1337, and a more diverse set of sequences ranging in overall sequence identities from 81.11% to 12.22%, which is closer to the spread for the CB1 sequences. These sequences are aligned, and the hydrophobic values for each amino acid are averaged over all 1337 sequences to produce an average hydrophobic value for each residue (**Figure 6.2**). The human CB2 sequence has seven hydrophobic regions corresponding to the seven alpha-helical TM regions.

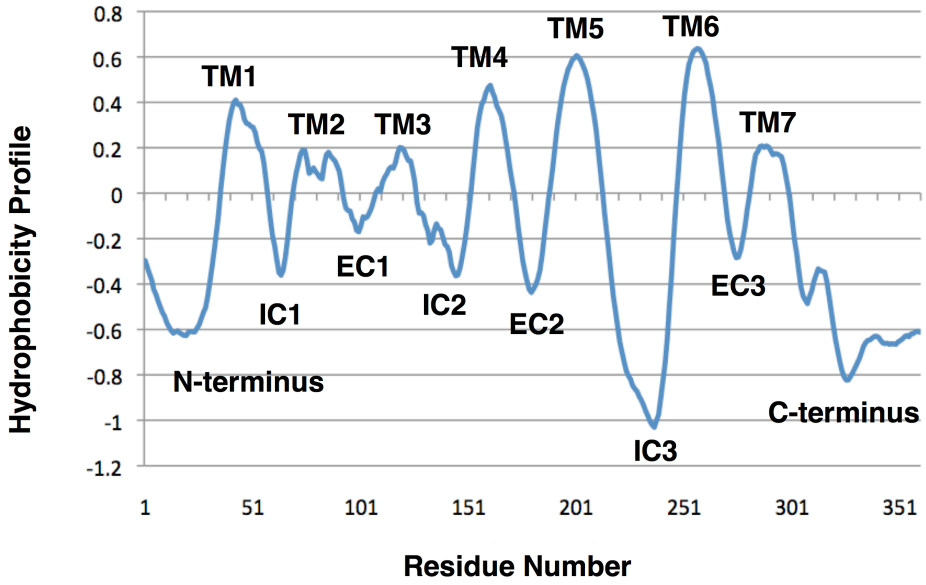


Figure 6.2 CB2 hydrophobicity plot. The seven TMs, three intracellular loops (ICs), three extracellular loops (ECs), N-terminus, and C-terminus regions are indicated.

Since the alpha-helical regions can extend beyond the hydrophobic lipid bilayer, the TMs provided by PredicTM are augmented with the SSP consensus estimate. Using prior knowledge from our predicted CB1 structures, we extended TMs 2, 4, and 6 to include residues that engaged in salt bridges and played an important role in stabilizing the CB1 receptor.

Optimizing CB2's TMs

The program OptHelix produces two types of helices based on different selection methods. The helices are selected from the MD simulations according to the minimum potential energy (mineng) structure and the structure closest in RMSD to the average structure of the simulation (minrmsd). The resulting minrmsd OptHelix helices are shown below in **Figure 6.3** with the corresponding CB1 helices. Six of the seven helices are very similar to each other in that their backbone root mean-squared displacement (RMSD) is less than 0.70 Å. Helix 6 has a substantial RMSD of 1.18 Å because CB1 has a glycine residue (rather than CB2's phenylalanine) adjacent to the proline residue in CWXP motif resulting in a differently kinked helix. Thus, small differences between the sequences of the receptors cause significant differences in the receptors' structures.

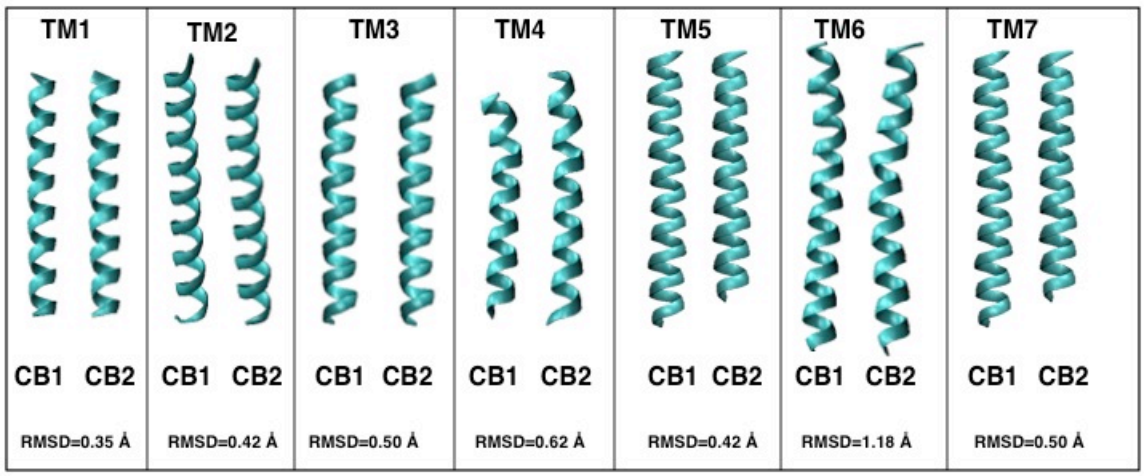


Figure 6.3 Comparison of the minrmsd helices generated by OptHelix for the CB1 and CB2 receptors.

Constructing the CB2 helical bundle

Since the CB1 and CB2 receptors are in the same family, we anticipate that they will be structurally similar, and we can use our predicted structure for the CB1 WT receptor, which has been supported by subsequent experiments (33; 34). We added the CB1 WT receptor to the available crystalized GPCR templates. The CB1 structure was the third lowest in energy, but it was chosen because it best represented the consensus of polar interactions present in the majority of the top ten lowest energy WT CB1 structures (for more information, see *Chapter III*). Since the CB1 structure prediction was done, many more GPCRs have been crystallized including the human lipid sphingosine 1-phosphate receptor type 1 (hS1P1) (PDB ID: 3V2Y) (40) and human dopamine D3 receptor (hDD3R) (PDB ID: 3PBL) (41) that are more closely related to the cannabinoid receptors (**Table 6.1A**). To date, the hS1P1 receptor has the highest TM sequence

identity with respect to CB2 of all the crystallized GPCRS at 35.92%. The inactive turkey β_1 adrenergic receptor (inactTβ_1AR) (chain B, PDB ID: 2VT4) (42) and the human β_2 adrenergic receptor (inactHβ_2AR) (PDB ID: 2RH1) (43) are also used because of their large TM sequence identities, 31.21% and 24.44% respectively, and because of the previous success we had using them for our CB1 structure predictions. These four structures (WT CB1, hS1P1, inactTβ_1AR, inactHβ_2AR) are the templates that the CB2 WT OptHelix helices are aligned to and are the starting points for the homology models.

Table 6.1 Constructing the WT, T3.46A, T3.46I receptors using first-based principles. **(A)** Comparison to the human CB2 receptor sequence of the overall sequence identity and TM sequence identity of human CB1 and crystallized GPCR sequences including hS1P1, Tβ_1AR, hDD3R, HA$_{2A}$AR, and Hβ_2AR.
(B) Energy comparison for CB1 receptors with OptHelix helices and homologized helices. Optimized eta (η, rotation) angles from the ComBiHelix analysis for the WT receptors with various helices: OptHelix helices aligned to the predicted CB1 model (33; 34), homologized CB1 helices, homologized inactTβ_1AR (PDB ID: 2VT4) (42) helices, homologized hS1P1 (PDB ID: 3V2Y) (40) helices, homologized inactHβ_2AR (PDB ID: 2RH1) (44) helices.

A.

GPCR Name	Overall Sequence Identity (%)	TM Sequence Identity (%)
Human CB2 Receptor	100.00	100.00
Human CB1 Receptor	41.11	54.56
Human S1P1 Receptor	21.11	35.92
Turkey Beta 1 Adrenergic Receptor	21.94	31.21
Human D3 Dopamine Receptor	15.83	27.43
Human A_{2A} Adenosine Receptor	17.50	25.66
Human Beta 2 Adrenergic Receptor	16.11	24.44

B.

	η (°)							TotalE
Method	**H1**	**H2**	**H3**	**H4**	**H5**	**H6**	**H7**	**kcal/mol**
OptHelix helices with CB1 template	-120	180	-30	30	90	-120	-60	-41.4
OptHelix helices with inactTβ_1AR template	0	90	30	120	30	60	-120	-5.3
Homologized CB1 Helices	-120	0	0	30	0	30	-60	52.3
Homologized inactTβ_1AR Helices	0	0	0	0	0	60	150	258.4
Homologized hS1P1 Helices	0	0	0	30	-90	90	-60	271.1
Homologized inactHβ_2AR Helices	0	0	0	0	-150	90	150	293.0

For the OptHelix helical bundles, there are two ways to align the helices to the template—where the horizontal plane bisects the helical bundle at the raw midpoint of the respective helices' hydrophobic region (the rawmid center) or at the residue splitting the area of the hydrophobic region (the area center). Thus for each of the four templates, there are four starting bundles generated [two types of helices (minrmsd, mineng) and two types of alignments (rawmid, area)]. Only one homologized helical bundle is created for each of the four templates since the helical shape and alignment is provided by the crystal structure.

Energetic comparison of templates

During the ComBiHelix stage, we identified the lowest total charged energy structure out of the 1000 structures that were constructed based on the BiHelix calculations. **Table 6.1B** shows the corresponding structures for all the homologized bundles plus the two lowest energy structures for the OptHelix helices. Since all of the structures contain the same number of atoms, we can compare their total charged energies after minimization. Much like the CB1 case, the CB2 structures from homologized crystal helices are much worse in energy than those predicted from the OptHelix helices or the homologized CB1 helices. Since the lowest energy bundles produced by the crystal homologized helices are ~200 kcal/mol and ~300 kcal/mol higher in energy than the CB1 homologized helices and the CB2 OptHelix helices, they were not used for the subsequent SuperBiHelix and SuperComBiHelix steps. The three structures that were the starting points for the subsequent local sampling were: the CB1 homologized helices, the OptHelix minrmsd helices aligned to the CB1 template with rawmid hydrophobic centers, and the OptHelix minrmsd helices aligned to the inact$T\beta_1AR$ template with area hydrophobic centers. SuperBiHelix testing entailed sampling ±10º in 10º increments in the theta (θ, tilt) angles, ±30º in 15º increments in the phi (φ, sweep) and eta (η, rotation) angles with the exception of the rotation angle in TM3 of the OptHelix helices aligned to CB1, which was sampled from -45º to +30º in 15º increments. The 2000 lowest energy structures predicted by SuperBiHelix were constructed as complete bundles in SuperComBiHelix.

Single point mutant structure prediction

To predict the T3.46A and T.346I mutants, we closely followed the procedure described above for the CB2 WT receptor. Starting with the OptHelix helices, the alanine or isoleucine residue replaced T.346 with the SCREAM program. The T3.46I mutant's OptHelix helices were aligned to all four templates with all helical and alignment combinations, while those belonging to the T3.46A mutant were aligned to the CB1 template with minrmsd helices at the rawmid center and

the inactTβ_1AR template with minrmsd helices at the area center. T3.46A and T3.46I were both constructed with the homologized helices from WT CB1 and hS1P1. For the T3.46A and T3.46I mutant, SuperBiHelix and SuperComBiHelix were performed on the lowest total charged energy structures from the OptHelix helices aligned to the WT CB1 template and the homologized WT CB1 helices using the same sampling angles described above.

RESULTS

Predictions of the stable structural ensembles for CB2 WT and mutant T3.46A and T3.46I receptors

As in the CB1 study, we used the GEnSeMBLE method to generate trillions of potential CB2 structures and identify ten low energy conformations that best represent the WT poses. For the CB2 WT conformation, seven of the top ten poses were generated from OptHelix helices aligned to the CB1 WT template (**Table 6.2**). The other three are generated from the CB1 homology helices. Similarly, for the T3.46I mutant, eight of the top ten poses resulted from CB2 OptHelix helices as opposed to two poses from the CB1 WT homologized helices. All of the top ten poses for the T3.46A mutant used the CB2 OptHelix helices.

The average energy ranges for the top ten poses for the three CB2 receptors are:

- T3.46A: -261.2 kcal/mol to -249.1 kcal/mol,
- WT: -304.6 kcal/mol to -277.4 kcal/mol,
- T3.46I: -272.5 kcal/mol to -247.9 kcal/mol.

And the backbone RMSD, with respect to the lowest energy structure, ranges are:

- T3.46A: 0.3 Å to 2.6 Å,
- WT: 0.4 Å to 2.9 Å,
- T3.46A: 0.4 Å to 2.9 Å.

The range for the RMSD values may seem large, but the top ten structures according to average energy rank were constructed from two types of helices—one type with the OptHelix program and another type constructed from homologized helices. The range of RMSD values for the CB2 WT OptHelix helices are 0.4 to 1.7 Å, all smaller than the best GPCR crystal resolution to date, which means that these structures are similar. Also, the RMSD range for the T3.46I structures from OptHelix helices is very small—0.5 to 0.8 Å. Thus, for these two receptors, even slightly different helical shapes impacted the overall structure of the TM bundles, whereas the structures generated with OptHelix helices are very similar.

Table 6.2 Comparison of the conformations of the low energy structures for the three CB2 receptors T3.46A, WT, and T3.46I. Methodology, theta (θ, tilt), phi (φ, sweep), eta (η, rotation) angles, average energy, and the backbone RMSD (Å) with respect to the lowest energy conformation for each of the ten structures are shown below. The poses were chosen according to their average energy rank, the average of the charged total energy rank, neutralized total energy rank, charged interhelical energy rank, and neutralized interhelical energy rank.

Name	Methodology	θ (°)							φ (°)							η (°)							Ave E	RMSD
		TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	(kcal/mol)	Å
T3.46A-1	WT CB1-rawmid-minrmsd	-10	-10	0	10	0	-10	0	0	30	30	0	0	0	-30	30	-30	30	-15	30	-15	-30	-255.5	0.0
T3.46A-2	WT CB1-rawmid-minrmsd	-10	-10	0	10	-10	-10	0	0	30	30	0	0	15	-30	30	-30	30	-15	0	-30	-30	-261.2	1.1
T3.46A-3	WT CB1-rawmid-minrmsd	0	-10	10	0	10	0	0	0	0	15	-30	-30	0	-15	30	-15	0	0	30	0	-30	-254.5	2.6
T3.46A-4	WT CB1-rawmid-minrmsd	-10	-10	0	10	0	-10	0	0	15	30	0	0	0	-30	30	-30	30	-15	30	-15	-30	-253.4	0.3
T3.46A-5	WT CB1-rawmid-minrmsd	0	-10	10	0	10	0	0	0	-15	15	-30	-30	0	-15	30	-15	0	0	30	0	-30	-254.5	2.6
T3.46A-6	WT CB1-rawmid-minrmsd	0	-10	10	0	10	0	0	0	15	15	-30	-30	0	-15	30	-15	0	0	30	0	-30	-254.5	2.6
T3.46A-7	WT CB1-rawmid-minrmsd	0	-10	10	0	10	0	0	0	-30	15	-30	-30	0	-15	30	-15	0	0	30	0	-30	-254.9	2.6
T3.46A-8	WT CB1-rawmid-minrmsd	0	-10	10	0	10	0	0	0	30	15	-30	-30	0	-15	30	-15	0	0	30	0	-30	-254.5	2.6
T3.46A-9	WT CB1-rawmid-minrmsd	-10	-10	0	10	-10	-10	0	0	15	30	0	0	15	-30	30	-30	30	-15	0	-30	-30	-257.5	1.1
T3.46A-10	WT CB1-rawmid-minrmsd	-10	-10	0	10	0	-10	0	0	30	30	0	0	15	-30	30	-30	30	-15	30	-15	-30	-249.1	0.3
WT-1	WT CB1-rawmid-minrmsd	0	0	0	0	0	0	0	0	0	-15	0	-15	30	0	0	0	0	0	0	-15	-15	-300.5	0.0
WT-2	WT CB1-rawmid-minrmsd	0	0	0	0	0	0	0	15	0	-15	0	-15	30	-15	0	0	0	0	0	-15	-15	-304.6	0.5
WT-3	CB1-Homology	10	0	0	0	0	-10	0	0	0	-15	15	0	15	15	0	0	0	0	-30	15	0	-291.5	3.0
WT-4	WT CB1-rawmid-minrmsd	0	0	0	0	0	0	0	15	0	-15	0	-15	30	-15	0	0	15	0	0	-15	-15	-286.3	0.6
WT-5	CB1-Homology	0	0	0	0	0	-10	0	-15	15	-15	15	0	15	15	0	0	0	0	-30	15	0	-282.9	2.9
WT-6	WT CB1-rawmid-minrmsd	0	0	0	0	0	0	0	15	0	-15	0	-15	30	-15	0	0	0	-15	0	-15	-15	-286.2	0.6
WT-7	WT CB1-rawmid-minrmsd	0	-10	-10	0	-10	0	0	-15	30	15	0	15	15	-15	15	-15	-15	0	-15	-15	-15	-279.0	1.7
WT-8	WT CB1-rawmid-minrmsd	0	-10	-10	0	-10	0	0	-15	15	15	0	15	15	-15	15	-15	-15	0	-15	-15	-15	-278.9	1.7
WT-9	WT CB1-rawmid-minrmsd	0	0	0	0	0	0	0	0	0	-15	0	-15	30	-15	0	0	15	0	0	-15	-15	-277.4	0.5
WT-10	CB1-Homology	0	0	0	0	0	-10	0	-15	0	-15	15	0	15	15	0	0	0	0	-30	15	0	-280.8	3.0
T3.46I-1	WT CB1-rawmid-minrmsd	0	0	0	0	0	-10	-10	15	-15	0	0	-15	0	0	0	0	15	0	-15	-15	-30	-272.5	0.0
T3.46I-2	WT CB1-rawmid-minrmsd	0	0	0	0	0	-10	-10	0	-15	0	0	-15	0	0	15	0	15	0	-15	-15	-30	-265.4	0.5
T3.46I-3	WT CB1-rawmid-minrmsd	0	0	0	0	0	-10	-10	0	-15	0	0	-15	0	0	0	0	15	0	-15	-15	-30	-265.6	0.5
T3.46I-4	WT CB1-rawmid-minrmsd	0	0	0	0	0	-10	-10	0	-15	0	0	0	-30	0	0	0	15	-15	0	-15	-30	-257.5	0.8
T3.46I-5	WT CB1-rawmid-minrmsd	0	0	0	0	0	-10	-10	15	-15	0	0	0	-30	0	0	0	15	0	0	-15	-30	-255.8	0.6
T3.46I-6	WT CB1-rawmid-minrmsd	0	0	0	0	0	-10	-10	0	-15	0	0	0	-30	0	0	0	15	0	0	-15	-30	-254.8	0.7
T3.46I-7	WT CB1-rawmid-minrmsd	0	0	0	0	0	-10	-10	0	-15	0	0	0	-30	0	15	0	15	0	0	-15	-30	-256.3	0.8
T3.46I-8	CB1-Homology	-10	-10	10	0	10	-10	-10	-15	-15	15	-30	-15	-15	-30	-15	-15	15	-15	0	15	-30	-250.2	2.9
T3.46I-9	CB1-Homology	-10	-10	10	0	10	-10	-10	-15	0	15	-30	-15	-15	-30	-15	-15	15	-15	0	15	-30	-249.8	2.9
T3.46I-10	WT CB1-rawmid-minrmsd	0	0	0	0	0	-10	-10	0	-15	15	0	0	-15	0	15	0	30	0	-15	-15	-30	-247.9	0.7

Intermolecular interactions in the constitutively active CB2 WT

Our predicted structure of the CB2 WT receptor contains two salt bridges characteristic of other class A GPCRs, D3.49+R4.41 and R3.50+D6.30 (**Figure 6.4A**). (Throughout this chapter, we will indicate a salt bridge with a '+' sign and a hydrogen bond with a '-'.) These two salt bridges are also observed in our predicted CB1 structures (**Figure 6.4B**). The former salt bridge is similar to one observed in the crystal structure of hCXCR4 (D3.49+K4.37) (45) and the one that may form in hDD3R (the arginine side chain was not resolved in the crystal) (41). As for the later salt bridge, it has been found in bRho (46-52), sRho (53; 54), hDD3R (41) and thermally stabilized inactHAA2AR (55) and inact$T\beta_1$AR (56) and also has been supported by experimental studies (57; 58). It is considered an ionic lock for class A GPCRs, in that it forms in the inactive state and needs to be broken for activation to occur. The other important salt bridges observed in the majority of low energy structures are K3.22+D5.38 and E1.49+R7.56. Surprisingly, the 1-2-7 and 2-3-4 hydrogen bond networks that are largely conserved in the class A GPCRs are absent from our CB2 structures because of the rotation of TMs 2 and 7. However, the 1-2-7 hydrogen bond network was absent in the CB1 WT receptor as well because of TM7's significant rotation. Surprisingly, T3.46 and S2.45 do not form hydrogen bonds in the CB2 WT receptor, as we had predicted for the CB1 receptor, even though both residues are conserved. TM2 is rotated so that

S2.45 faces TM1 not TM3. CB2 does contain a hydrogen bond between TM3 and TM4 via an interaction between T3.37 and S4.57.

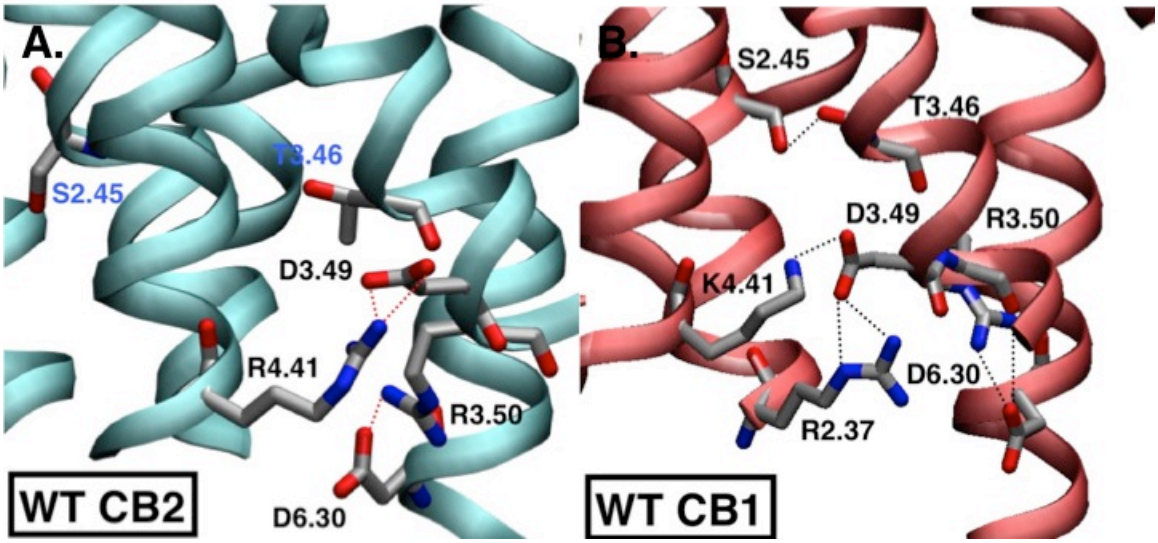


Figure 6.4 Comparison of the intracellular ends of the **(A)** CB2 WT (cyan) and **(B)** CB1 (red). Salt bridges in the CB2 WT structure that are conserved in the CB1 WT receptor are represented by dotted red lines. Residues that participate in polar contacts in CB1 WT but not in CB2 are labeled in blue. A pose that most accurately represents the consensus of low-lying energy structures, in that it had the same polar contacts that the majority of the ten structures, was selected for each receptor and shown above. CB2 WT is represented by the second lowest energy conformation, and the CB1 WT is represented by the third lowest energy conformation. These respective structures will be used throughout the rest of this chapter.

Intermolecular interactions in the CB2 T3.46A receptor

Without any prior knowledge of agonist or G-protein binding to the mutant T3.46A receptor, we predicted the structures using both CB1 WT homologized helices as well as CB2 OptHelix helices constructed from canonical poly-alanine helices. According to GTPγS assays and ligand binding assays, this threonine to alanine mutation caused the CB1 receptor to become fully inactive (33; 34; 36; 37). According to **Table 6.2**, all of the lowest average energy T3.46A structures were constructed with the CB2 OptHelix helices. By analyzing the results of the top ten structures, we see a lot of variability in the polar contact patterns. There are 10 hydrogen bonds that appear in five of the ten top structures. Another three hydrogen bonds appear in at least one of the ten structures. Many conformations have the D3.49+R4.41 salt bridge, which was seen in the CB2 WT receptor, and the D5.38+K7.33 salt bridge (**Figure 6.5**). Surprisingly, in the CB2 T3.46A receptor, there are no R3.50+D6.30 and R2.37+D6.30 salt bridges that were present in the CB1 T3.46A receptor. **Figure 6.5B** shows that the important residues that participate in the TM3+TM6 ionic lock are not actively forming interhelical interactions. Also, as in the CB2 WT receptor, the typical 1-2-7 and 2-3-4 hydrogen bond networks are not present despite having all the participating residues conserved in the receptor. However, there are non-

traditional 1-2-7 and 2-3-4 hydrogen bonds. For example, instead of interactions between N1.50, D2.50, and N7.49, there are hydrogen bonds between K1.32-N2.63 and E1.49-Y7.47 (the Y of the NPXXY motif). As for the 2-3-4 hydrogen bond network, instead of the polar contacts between S2.45, S3.42, and W4.50, there are hydrogen bonds between T3.37 and S4.57 as well as S2.60 and K3.28. None of these interactions are observed in the CB1 T3.46A receptor.

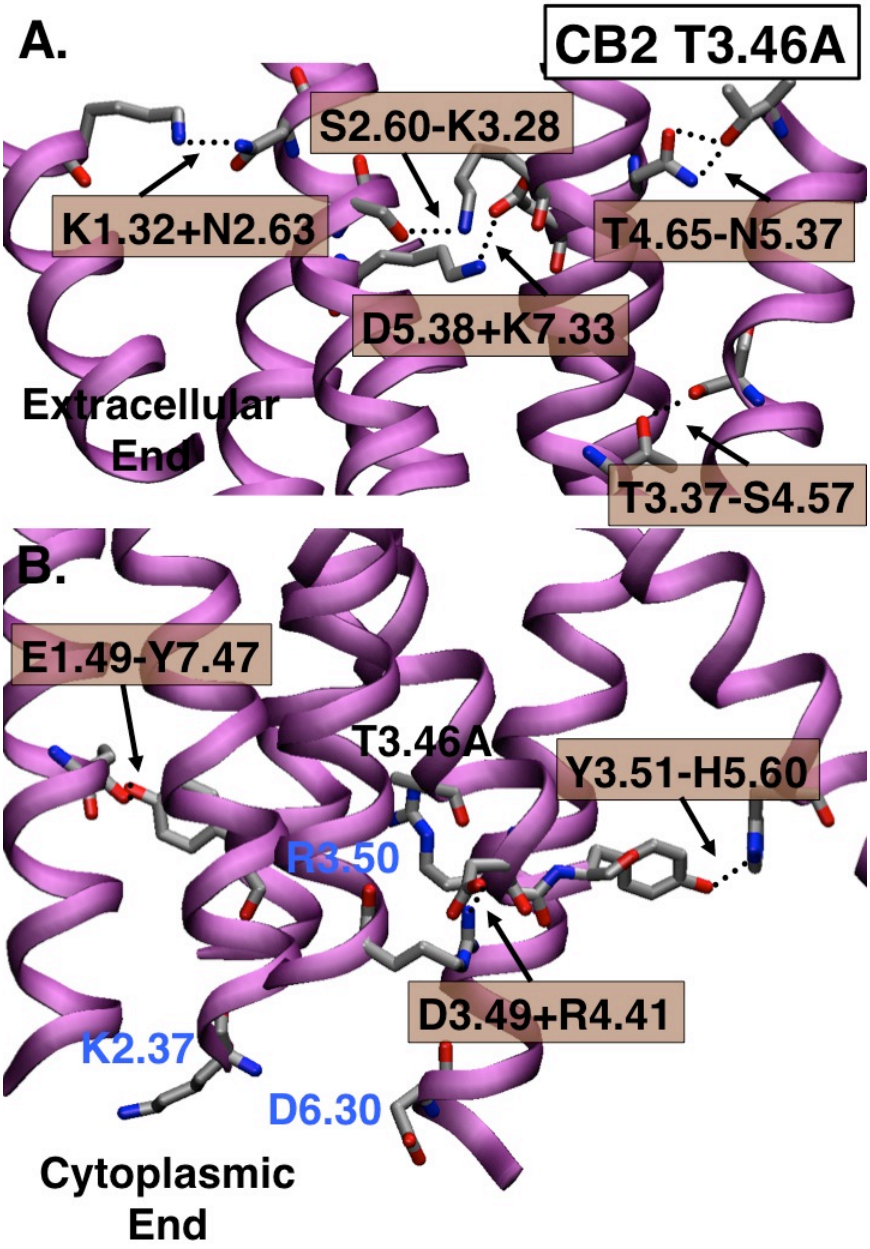


Figure 6.5 Salt bridges and hydrogen bonds on the **(A)** extracellular end and the **(B)** intracellular end of the predicted structure of the CB2 T3.46A receptor. Residues that participate in polar contacts in CB1 T3.46A, but not in CB2 T3.46A, are labeled in blue. The structure shown (sixth lowest in energy) best represents the consensus of the polar contacts in the top ten predictions for the CB2 T3.46A structure. This structure will be used when discussing T3.46A throughout the rest of this chapter.

Intermolecular interactions in CB2 T3.46I

Using a methodological approach similar to the one described above for the T3.46A receptor, we predicted the structure of the CB2 T3.46I receptor. The corresponding mutant for the CB1 receptor had been proven experimentally to be highly constitutively active (33; 34; 36; 37). We do not observe the D2.63+K3.28, R2.37+D3.49, and R5.71+D6.30 salt bridges in the CB2 structures that are present in CB1 T3.46I (**Figure 6.6**). In fact, other than D3.49, none of these residues consistently form interhelical polar contacts in the CB2 version of the receptor. However, CB2 T3.46I does contain K3.22+D5.38 and D3.49+R4.41 salt bridges, both of which are observed in the predicted structures of the CB2 WT receptor. The D3.49+R4.41 salt bridge is present both in the CB2 WT and T3.46A receptors as well as the full inactive CB1 T3.46A and weakly constitutively active CB1 WT receptor, but not the highly constitutively active CB1 T3.46I receptor. In fact, the K3.22+D5.38 salt bridge is the only polar interaction in CB2 T3.46I that does not occur between adjacent helices. For example, the CB2 T3.46I receptor has hydrogen bonds between D2.50 and S3.39, Y5.39 and H6.57, and W6.48 and S7.46. It is notable that the only polar connection between non-adjacent helices is on the extracellular end of the helical bundle. The CB2 T3.46I receptor, much like the CB1 T3.46I receptor, has fewer hydrogen bonds than its corresponding WT and T3.46A receptors. Also, this isoleucine mutant, like the other cannabinoid receptors and their various mutants, does not have the classical 1-2-7 or 2-3-4 hydrogen bond networks.

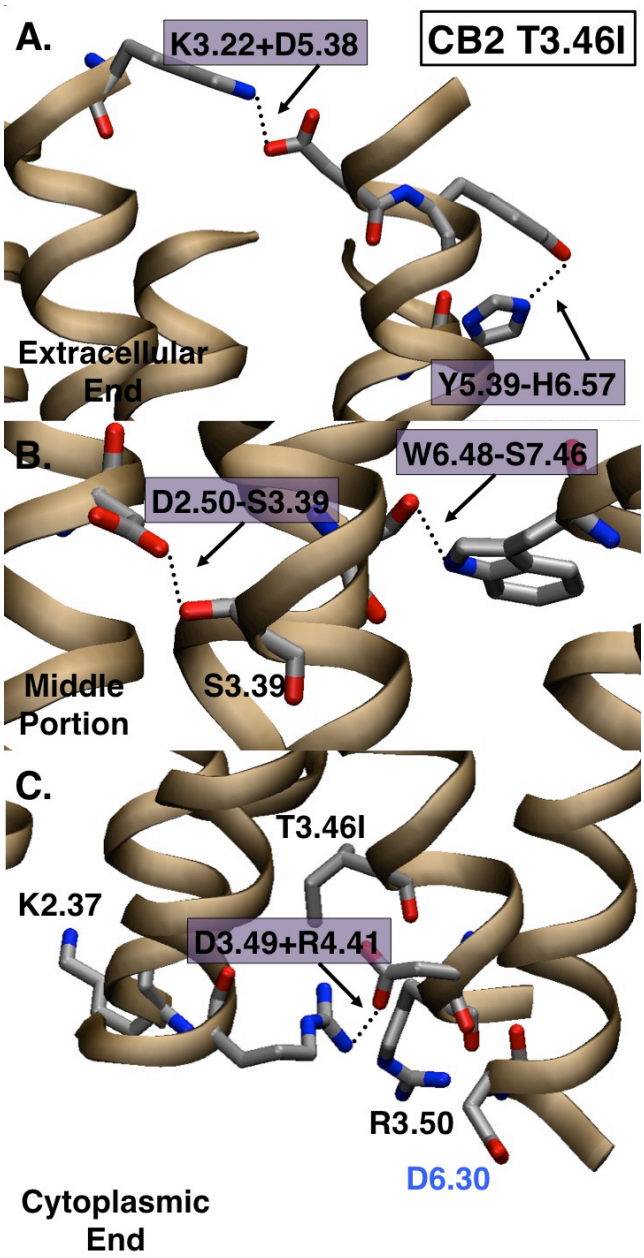


Figure 6.6 Salt bridges and hydrogen bonds on the **(A)** extracellular end, **(B)** the middle portion, and the **(C)** cytoplasmic end of the predicted structure of the CB2 T3.46I receptor. The structure shown (fourth lowest in energy) best represents the consensus of the polar contacts in the top ten predictions for the CB2 T3.46I structure. Residues that participated in polar contacts in CB1 T3.46I are labeled in blue. This structure will be used when discussing T3.46I throughout the rest of this chapter.

Salt bridge patterns differ across the three CB2 receptors

Table 6.3 shows a comparison of the consensus of polar interactions observed in the three CB2 receptors. The CB2 WT receptor has the classic R3.50+D6.30 that is believed to be an ionic lock that stabilizes the inactive form in class A GPCRs. CB2 site-directed mutagenesis data, however, indicates that the TM3+TM6 ionic lock may not be present or necessary for stabilizing the

inactive CB2 structure. Two separate studies show that the R3.50A mutation in CB2 does not affect agonist binding (59; 60), which suggests that despite having a broken TM3+TM6 coupling, the mutant receptor remains inactive. Hence, we predict that a TM2+TM6 ionic lock would be present to stabilize the inactive conformation much like the way it does in the fully inactive CB1 T3.46A mutant.

Table 6.3 Comparison of consensus salt bridge patterns for each of the three CB2 receptors (T3.46A, WT, and T3.46I).

Receptor	K2.37 + D6.30 ionic lock	R3.50 + D6.30 ionic lock	D3.49 + R4.41 salt bridge	K3.22 + D5.38 salt bridge	D5.38 + K7.33 salt bridge	E1.49 + Y7.52/ R7.56 salt bridge	T3.37 - S4.53/ S4.57 hydrogen bond	D2.50 - S3.39 hydrogen bond
T3.46A	No	No	Yes	No	Yes	Yes	Yes	No
WT	No	Yes	Yes	Yes	No	Yes	Yes	No
T3.46I	No	No	Yes	Yes	No	No	No	Yes

All of the receptors are in agreement in that they all are expected to contain the D3.49+R4.41 salt bridge that was observed in fully inactive and weakly constitutively active CB1 mutant and WT receptors. This is the only polar connection observed in both the CB1 and CB2 receptors. The CB2 WT and T3.46I receptors contain a K3.22+D5.38 extracellular salt bridge, whereas in the CB2 T3.46A mutant, D5.38 interacts with K7.33 instead. While the traditional 1-2-7 and 2-3-4 hydrogen bonds are absent in these CB2 structures, there are unusual connections. TM1 in the CB2 WT receptor is anchored via a salt bridge between E1.49 and R7.56. This salt bridge is present in half of the CB2 T3.46A structures, but the other half of the structures contain a analogous E1.49-Y7.52 hydrogen bond. Similarly, CB2 WT has a T3.37-S4.57 hydrogen bond as do half of the CB2 T3.46A structures. The other half have a T3.37-S4.53 hydrogen bond instead. The CB2 T3.46I mutant has a unique D2.50-S3.39 hydrogen bond not observed in the other receptors. Usually, D2.50 interacts with N1.50 and N7.49 in the crystallized class A GPCRs. The hS1P1 receptor, the closest related crystallized GPCR to CB2, has a hydrogen bond between D2.50 and S3.39 because their oxygen-oxygen heteroatom distance is 3.12 Å. This structural feature indicates that TM2 is rotated so that D2.50 can interact with TM3. However, D2.50 simultaneously forms a hydrogen bond with N7.49 (40), which is impossible in our predicted CB2 WT structure because of TM2's significant rotation away from the TM 1-2-7 pocket and because of TM7's orientation.

Ligand binding to CB1/2 WT, T3.46A, and T3.46I

Our collaborators, Professor Debra Kendall and her lab at the University of Connecticut, have analyzed the cannabinoid ligands' binding affinity to the CB2 and CB1 WT receptors and two mutants, T3.46A and T3.46I (**Table 6.4**). Previously, using GTPγS assays, Ahn and co-workers showed that for the CB1 receptor, T3.46A is fully inactive and T3.46I is highly constitutively active (33; 34). The ligand binding affinities, values of K_i, agree with the activity trends shown in the GTPγS assays in that the Rimonabant antagonist/inverse agonist has a three-fold stronger affinity for the inactive T3.46A mutant, but has a 25-fold weaker affinity for the active T3.46I with respect to the CB1 WT receptor. Conversely, the agonist WIN55212-2 has a three-fold improved binding affinity for the active T3.46I mutant, but a 25-fold weakened binding affinity for the inactive T3.46A mutant. While there are no GTPγS assays done for the CB2 WT and mutant receptors, we do have data for the agonist binding affinities. The K_i values of CP55940 for the CB2 WT and T3.46A mutant are 1.39 nM and 1.42 nM respectively. These data indicate that cannabinoid agonists have the same attraction for both the mutant CB2 T3.46A and the WT receptors.

Table 6.4 Ligand binding properties of CB1 and CB2 T3.46A, WT, and T3.46I receptors. The competitor ligand was [^{3}H]CP55940 or [^{3}H]Rimonabant (indicated by an *).

Cannabinoid Receptor	Ligand Type	Ligand Name	T3.46A K_i (nM)	WT K_i (nM)	T3.46I K_i (nM)	K_i Ratio T3.46A:WT	K_i Ratio WT:T3.46I
CB1 (36; 37)	Anta-gonist	Rimonabant	2.2*	7.2*/7.5	190	1:3	1:25
CB1 (36; 37)	Agonist	WIN55212-2	1440*	68*/73	28.3	21:1	3:1
CB2	Anta-gonist	TBD	TBD	TBD	TBD	TBD	TBD
CB2	Agonist	CP55940	1.39	1.42	0.23	1:1	6:1

As for the CB2 T3.46I mutant, there are observable differences in the agonists' binding affinity for it versus the WT receptor. CP55940 has a K_i value of 0.23 nM for the T3.46I receptor, which is a six-fold stronger binding affinity than for the WT receptor. These data follow the patterns for the CB1 WT and T3.46I receptor in that the T3.46I receptor has slightly stronger affinities for agonists.

DISCUSSION

The WT CB2 structures differ from their CB1 counterparts

We anticipated that the two members of the cannabinoid family would share many structural similarities because of their high percentage of conserved amino acid identity within the TMs,

54.56%, but instead there are substantial differences between their predicted structures. After testing a variety of crystal templates, we produced substantially lower energy structures using the previously predicted CB1 template as we had anticipated. Many of the residues that stabilized CB1 and its various mutants are conserved in CB2. However, it seems that the residues that are not conserved play a major role in influencing the structural differences. Of the top ten lowest energy structures, the vast majority of them used the CB2 OptHelix helices, developed from canonical poly-alanine helices modified with SCREAM, MD, and minimization, rather than the mutated homology helices from the CB1 WT predicted structure. OptHelix helices are structural optimized for the specific target sequence, so the resulting TM bundles built using OptHelix helices or CB1 homologized helices have large structural differences. The CB1 structure was also constructed from OptHelix helices rather than homologized helices. Unlike CB2, the lowest energy CB1 structures were built with mineng helices, not minrmsd helices. The RMSD for the two minrmsd helices 6 is 1.18 Å, but the RMSD between CB2 minrmsd helix 6 and CB1 mineng helix 6 is much less at 0.51 Å, so the lowest energy CB2 structures prefer a certain conformation of helix 6 that is captured by the CB2 minrmsd OptHelix helices or the homologized CB1 mineng OptHelix helices.

The biggest difference between the CB1 and CB2 WT predicted structures is the orientation of TM2. In CB1 WT, we observed two salt bridges between helices 2 and 3 that we proved with mutations to be important for stabilizing the inactive or active conformations (33; 34). For example, R2.37 in CB1 participates in a TM2+TM6 coupling in the fully inactive mutant receptors such as CB1 T3.46A (**Figure 6.7A**). This salt bridge breaks to form a new one between R2.37 and D3.49 in the constitutively active mutants. It was expected that K2.37 in CB2 would play a similar role. Likewise, D2.63 forms a salt bridge with K3.28 in weakly or highly constitutively active receptors like CB1 L3.43A (**Figure 6.7C**). By mutating the aspartic acid to an alanine, the active mutant became fully inactive. It was also expected that Q2.63 in CB2 would interact with the conserved K3.28 residue. Also, residues S2.45 and T3.46 are conserved in both receptors. In the CB1 WT receptor, these two residues formed a hydrogen bond (**Figure 6.7E**), which might explain why the T3.46 residue is so important for the CB1 receptor structure, and it was anticipated to have a similar role in the CB2 receptor.

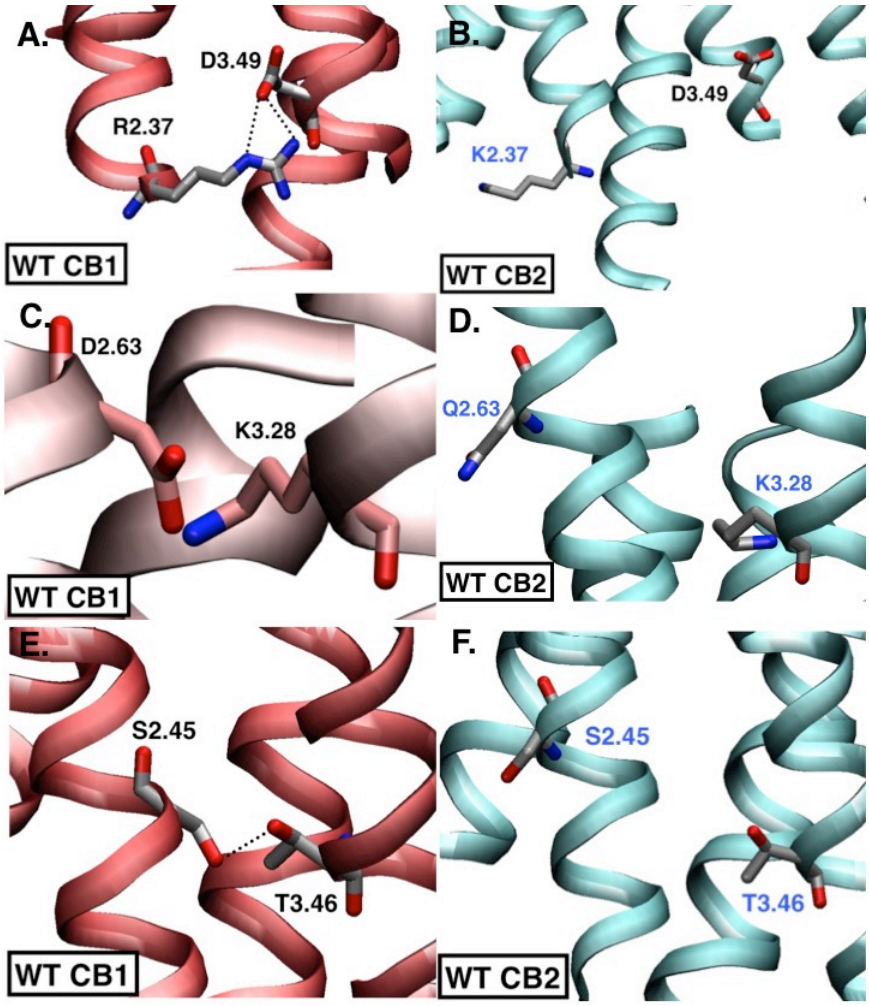


Figure 6.7 Comparison of CB1 (red) and CB2 (cyan) TM2 residue placement for **(A)** R/K2.37 and D3.49, **(B)** D/Q2.63 and K3.28, and **(C)** S2.45 and T3.46.

However, none of these contacts were present in the sister CB2 WT receptor (**Figures 6.7 B, D, and F**). It does not have any polar interactions between TMs 2 and 3, and the CB1 salt bridges and hydrogen bond networks cannot possibly form. TM2 is rotated almost 60° from where it is oriented in the CB1 WT receptor and is stabilized by two lysine residues, K2.37 and K1.32, which form hydrogen-bonds with backbone atoms on adjacent helices. The CB2 WT receptor's different orientation of TM2 suggests that there are significant structural differences, and thus functional differences, with respect to CB1.

There are other differences in the salt bridge patterns between the two types of cannabinoid receptors. A D6.58+K7.32 interaction forms in the majority of CB1 WT structures, but this salt bridge is not observed in CB2 WT because there is a serine instead of an aspartic acid at position 6.58 in the CB2 receptor. It does, however, have several salt bridges that are not shown in the

CB1 WT structures. For example, the CB2 WT receptor has an extracellular polar contact between K3.22 and D5.38. This salt bridge is not observed in CB1 because CB2's aspartic acid corresponds to a serine at position 5.38 in CB1, and R3.22 was not included in CB1's TM3 because it was predicted to be in the loop region. This TM3-TM5 interaction can form in CB1 during dynamic simulations. Furthermore, the CB2 WT receptor shows a strong preference for forming a salt bridge between TM1+TM7, E1.49+R7.56. This polar contact stabilizes the TM1 orientation, which was quite variable in the CB1 WT receptor. The R7.56 residue was not included in TM7 in CB1 WT because it was not predicted to be in the helical region. This salt bridge does form during dynamics provided that TM7 is in an orientation that makes the residue accessible. We observe the formation of this salt bridge during CB1 WT MD simulations. Perhaps this interaction is why the predicted structures of the cannabinoid receptors do not contain the typical 1-2-7 hydrogen bond networks—because TM1 and TM7 need to be oriented a specific way in order to form that salt bridge, which is not conducive to the characteristic class A GPCR hydrogen bond network. The only crystallized GPCR that contains the E1.49 residue is the hS1P1 receptor (40), but the side chain was not resolved (PDB ID: 3V2W), and the residue at 7.56 is a tyrosine, so E1.49 does not form an interhelical salt bridge in hS1P. Thus, hS1P1 does contain the expected 1-2-7 network. In general, MD simulations may prove that the two cannabinoid receptors are more similar than the loopless and static predicted structures indicate.

G2.48 may play an important role in the predicted WT CB2 structure

We propose that the G2.48 residue in CB2 is what causes TM2 to have an unexpected orientation. The TM3s of the CB1 and CB2 WT receptors are very similar to each other. TM2 has a greater variation in its sequence. **Figure 6.8A** shows CB2 WT's TM1 and TM2 with N1.50, D2.50, and G2.48 indicated. CB2 has a glycine at position 2.48 whereas CB1 has a much bulkier valine. We believe that this particular residue may play an important role in positioning TM2. The predicted distance between the nitrogen atom in the side chain and glycine's Cα atom is 3.25 Å. This is not enough to accommodate a side chain without clashing with N1.50. TM2 would have to rotate in order for the valine residue at position 2.48 to have enough space.

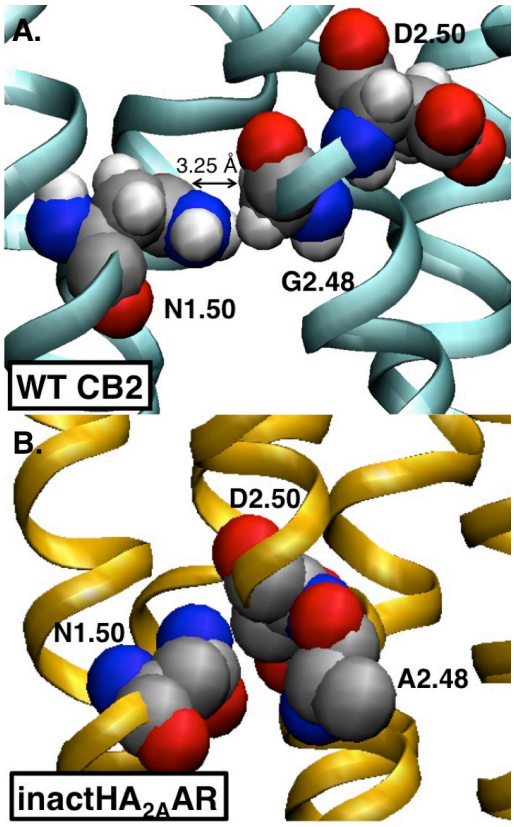


Figure 6.8 Orientation of residues that impact the TM2 orientation in **(A)** CB2 and **(B)** inactHA$_{2A}$AR (PDB ID: 3EML) (61)**.**

Figure 6.8B shows the positions of the corresponding residues from **Figure 6.8A** in the inactive human adenosine A_{2A} receptor (inactHA$_{2A}$AR) crystal structure (PDB ID: 3EML) (61). N1.50 is in relatively the same position, but D2.50 and A2.48 are roughly a quarter-turn from where they are in the CB2 WT receptor. In CB2, G2.48 faces TM1, and D2.50 faces TM3. In inactHA$_{2A}$AR, A2.48 is facing into the lipid bilayer, and D2.50 faces into the middle of the bundle to form hydrogen bonds with water molecules and other residues. And in inactHA$_{2A}$AR, unlike in CB2, TM2 is tilted so that N1.50 is not directly pointing at a residue. CB2's TM2 orientation would not be feasible with the inactHA$_{2A}$AR amino acid sequence. By replacing CB2's G2.48 with an alanine, we expect that the TM2 rotation will change significantly. The other option would be for TM1 to move to accommodate the larger alanine residue, but we do not anticipate this because in its current conformation, CB2 TM1 is stabilized by a salt bridge with TM7, (E1.49+R7.56).

Glycine at the 2.48 spot is found in very few GPCRs. Of the crystallized class A GPCRs, which all have a 1-2-7 network, hCXCR4, bRho, and hDD3R have a valine (like CB1), HA$_{A2}$AR has an alanine, Tβ_1AR and Hβ_2AR have a cysteine, and sRho has a phenylalanine residue at this

position. Of the 1559 related sequences found in BLAST, only the CB2 receptors and certain Mas-related GPCRs, gp162, olfactory receptors, ribosome maturation factors, and the Protease HtpX homologs have a glycine at this position. None of these proteins have been crystallized, so it is unknown if the 1-2-7 hydrogen network forms. In the latter two proteins listed above, the residues participating in the 1-2-7 network are not conserved, so the polar contacts must be different.

To confirm our prediction, it would be necessary to predict the structures of the CB2 G2.48V and CB2 G2.48V/T3.46A. We propose that this G2.48V mutation would force TM2 to change its orientation. The former mutant should resemble the CB1 WT receptor in that it should have a polar TM2+TM3 coupling. The latter mutant should compare with the fully inactive CB1 T3.46A mutant in that it should have a TM2+TM6 ionic lock. If these structural predictions are correct, then our collaborators will perform GTPγS binding assays to determine the degree of G protein binding and the respective activity levels. Thus, we expect that the G2.48V receptor has a higher level of GTPγS binding than the G2.48V/T3.46A double mutant, and thus has a higher level of constitutive activity than the G2.48V/T3.46A mutant, and that there will be a clear distinction between the two receptors' function and behavior. If the results do not change with the mutation of G2.48 to valine, then the residue at that position is unimportant in orienting TM2 and probably does not interact with N1.50. If our predictions are correct, then that would indicate the G2.48 residue plays a significant role in differentiating the structures and signaling.

The CB2 T3.46A mutant is constitutively active

We conducted structure prediction experiments that were performed independently and before we had experimental data to attest the functional state of the CB2 T3.46A and T3.46I mutants. In our predicted structures, the CB2 T3.46A receptor lacked a TM2+TM6 coupling much like the predicted structures of the CB1 WT and T3.46I receptors that have been experimentally proven to be constitutively active. The R3.50+D6.30 ionic lock is also absent in this CB2 T3.46A mutant as in the constitutively active CB1 T.346I receptor. Thus we conclude, based on our structural predictions, that CB2 T3.46A is activated, which is different from what was observed in the CB1 receptor.

Even though our predicted structure of CB2 T3.46A is lacking a TM3+TM6 ionic lock, it does not mean that the structure is in the highly constitutively active conformation as opposed to the weakly constitutively active one. In our designed CB1 mutants, T3.46A/R2.37A and

L3.43A/D2.63A/R2.37A, the predicted structures lacked this same salt bridge and are stabilized by the D3.49+K4.41 salt bridge, which is not present in the highly constitutively active mutants like T3.46I or L3.43A. Experimental GTPγS binding assays proved that these mutants are constitutively active, and not highly constitutively active, as we had predicted (33; 34).

Subsequent agonist binding assays (**Table 6.4**) support our predicted CB2 T3.46A structure. The agonist CP55940 has similar binding affinity for both the CB2 WT and T.346A receptors, which suggests that they are equally activated. While this preliminary data looks very promising, there are other experiments that would be necessary to inconclusively prove that both these receptors are constitutively active. First, the experiments with agonists need to be repeated, so there will be a margin of error for the agonists' binding affinities. Second, the experimental binding affinities of antagonists needs to be measured to all of these receptors. If they are constitutively active, we would expect the antagonists to have equal binding affinities to the mutant receptors with respect to the CB2 WT in the same way the agonists did. If the CB2 WT, T3.46A, and T3.46I receptors do not have equal activation levels, then the antagonists would bind strongest the receptor that is least active. For example, we anticipate that the antagonists will have larger binding affinities for the CB2 WT and T3.46A receptors than the T3.46I mutant because, based on the agonist binding data, we expect CB2 T3.46I to be more highly activated.

It is reasonable to assume that the agonists would prefer to bind to the more activated conformation. However, we cannot come to any definitive conclusions about activation levels until we have the results from the GTPγS assays. These assays will determine the level of GTPγS binding, an indicator of G protein binding, in absence of the ligand, which is a more accurate indicator of constitutive activity than ligand-induced activation. Based on the agonist binding affinities, the T3.46I mutant is expected to have a larger GTPγS binding affinity and thus, be more constitutively active, and the WT and T3.46A receptors will have weaker binding affinities than T3.46I and be equal to one another.

The CB2 T3.46I mutant is also constitutively active

We also predicted the structure of the CB2 T3.46I mutant and found that it too lacked the TM2+TM6 and TM3+TM6 couplings that are known to stabilize the inactive conformations of the CB1 receptor. However, it also lacks the R2.37+D3.49, D2.63+K3.28, and R5.71+D6.30 salt bridges that are present in the highly constitutively active CB1 T3.46I receptor. One of these, D2.63+K3.28, is believed to stabilize the constitutively active conformations (33). The

R5.71+D6.30 salt bridge is only observed in the constitutively active CB1 T.346I receptor, but similar ones are observed in the putatively active bOps receptor (62; 63), the constitutively active bRhod mutant (64), the meta II receptor (65), and perhaps a water-mediated one in the nanobody-bound act$H\beta_2AR$ (66). However, our predicted structure for the highly constitutively active L3.43A mutant lacks this salt bridge as well (33), but it did contain both TM2+TM3 couplings. Surprisingly, this CB2 T3.46I mutant contains the D3.49+R4.41 salt bridge, which is present in the fully inactive CB1 T3.46A mutant and other weakly constitutively active CB1 mutant and WT receptors (33). Similar salt bridges have been observed in antagonist-bound receptors (41; 45). Possibly, this contact stabilizes the inactive and weakly constitutively active receptors. The CB2 T3.46I receptor does have a K3.22+D5.38 salt bridge, which is the only contact in the majority of structures that is not between adjacent helices. It occurs at the extracellular end, so the intracellular end is spacious enough for potential proteins to bind. Despite the lack of agreement with the CB1 T3.46I structure, the CB2 T3.46I prediction appears constitutively active due to its lack of the two major ionic locks. Though surprisingly, of the top ten lowest energy structures, three of them had the R3.50+D6.30 ionic lock and two of them had the R2.37+D6.30 ionic lock, but these locks were never present in the same structure or the majority of conformations.

Our predicted structures agree with experimental results, in that agonist CP55940 has a six-fold improvement in binding affinity for the CB2 T3.46I mutant compared to the CB2 WT or T3.46A receptors, which is a strong indicator that T3.46I is constitutively active. However, there will not be any definite conclusions without more ligand binding assays and the GTPγS assays. Based on our structural predictions, which contain the D3.49+R4.41 salt bridge, but none of the TM2+TM3 or TM5+TM6 salt bridges, CB2 T3.46I is not expected to be highly constitutively active like its CB1 counterpart.

CB2 WT, T3.46A, and T3.46I structures are very different from one another, and the receptors may have different functional properties

Even though there is some agonist binding evidence to suggest that the CB2 WT, T3.46A, and T3.46I receptors are all constitutively active, they are structurally very different. The CB2 WT conformation has a conserved R3.50+D6.30 salt bridge absent in the T3.46A one, and they have a significant RMSD of 2.77 Å, which is larger than the value for active versus inactive crystallized GPCRs (43; 50; 61; 62; 65; 67-69). The backbone RMSD between the CB2 WT and T3.46I receptor is slightly smaller, at 2.28 Å. The two mutants have a significantly different RMSD, of

3.22 Å. The structural diversity could indicate functional differences for the three constitutively active receptors.

GPCRs have multiple active states as explained by the ensemble state theory (70-73) and crystallized Hβ_2AR conformations (43; 66; 69). As discussed earlier, the CB1 receptor is a prime example because of its ability to bind to multiple types of G proteins including G_i (74) and G_s (37; 75; 76) as well as the β arrestin proteins (64; 77-79). Our computational studies in *Chapter III* agree with this concept because we predict different conformations for the CB1 WT and L3.43A receptors and they have been experimentally proven to bind to different G proteins (33; 37; 74). Furthermore, previous experiments show that agonists affect G protein binding to the CB1 receptor and subsequent activation. For example, when CB1 binds to the G_i protein, agonist CP55940 had approximately the same level of activity as WIN55212-2. However, when binding to a G_s protein, the CP55940 had only 45% of the activity that WIN55212-2 produced. Possibly, the agonists stabilize different active conformations of CB1 that lead to different intracellular reactions with the G proteins and thus different activity levels (80). Perhaps CB2 experiences similar phenomena.

Based on the differences between the predicted structures, we anticipate that the CB2 WT and T3.46A receptors have different functions despite having the same binding affinity for the agonists. Much less is known about the CB2 signaling pathway than the CB1 one, but it is reasonable to believe that the receptor might have different conformations to interact with different types of proteins. For example, the CB2 WT receptor has been shown experimentally to bind to the G_i protein, but the CB2 T3.46A mutant may bind to the G_s or G_q proteins or perhaps one of the β arrestin proteins. However, these conjectures about the CB2 structures need to be validated with molecular biology experiments. Cyclic adenosine monophosphate (cAMP) assays will indicate which type of G protein the two mutant receptors bind to. These assays will show how well our structural predictions can indicate differences in function and GPCR behavior.

CONCLUSIONS

We predicted the structures of the CB2 WT plus two mutant receptors, T3.46A and T3.46I. Previous experiments showed that mutating this threonine residue as position 3.46 significantly changes the function of the CB1 receptor, and we were curious to see if this was true for its sister cannabinoid CB2 receptor. Surprisingly, despite the large sequence conservation between the

two receptors, the resulting WT structures are very different. TM2 is rotated so that none of the TM2-TM3 polar interactions observed in the CB1 receptor are present in the CB2 one. Furthermore, the CB2 T3.46A mutant, which was fully inactive in CB1, lacks the R2.37+D6.30 and R3.50+D6.30 ionic locks that prevent activation. We predict that this mutation does not have the same impact in CB2 as it does on CB1, and therefore, the T3.46A receptor is constitutively active. Similarly, the T3.46I mutant also lacks these important ionic locks, so it too resembles a constitutively active receptor. Our structure predictions are supported by agonist binding assays done by the Kendall lab. They show that agonists have equal binding affinities for the CB2 WT and T3.46A receptors and a slightly stronger affinity for the T3.46I mutant. However, antagonist binding and GTPγS assays need to be performed to conclusively prove these findings. In the meantime, we have designed a double mutant, CB2 G2.48V/T3.46A, that is predicted to contain the R2.37+D6.30 ionic lock and thus be fully inactive. After verifying our receptor structure predictions, we can subsequently design anti-obesity drugs and pain medication to interact with this CB2 receptor without resulting in psychological side effects.

REFERENCES

1. Howlett AC, Barth F, Bonner TI, Cabral G, Casellas P, Devane WA, Felder CC, Herkenham M, Mackie K, Martin BR, Mechoulam R, Pertwee RG (2002) International Union of Pharmacology. XXVII. Classification of cannabinoid receptors. Pharmacol Rev **54**:161-202.
2. Matsuda LA, Lolait SJ, Brownstein MJ, Young AC, Bonner TI (1990) Structure of a cannabinoid receptor and functional expression of the cloned DNA. Nature **346**:561-564.
3. Devane WA, Dysarz FA, Johnson MR, Melvin LS, Howlett AC (1988) Determination and characterization of a cannabinoid receptor in rat brain. Mol Pharmacol **34**:605-613.
4. Herkenham M, Lynn AB, Johnson MR, Melvin LS, Decosta BR, Rice KC (1991) Characterization and localization of cannabinoid receptors in rat brain: A quantitative *in vitro* autoradiographic study. J Neurosci **11**:563-583.
5. Herkenham M, Lynn AB, Little MD, Johnson MR, Melvin LS, Decosta BR, Rice KC (1990) Cannabinoid receptor localization in brain. Proc Natl Acad Sci USA **87**:1932-1936.
6. Jansen EM, Haycock DA, Ward SJ, Seybold VS (1992) Distribution of cannabinoid receptors in rat brain determined with aminoalkylindoles. Brain Res **575**:93-102.
7. Galiègue S, Mary S, Marchand J, Dussossoy D, Carrière D, Carayon P, Bouaboula M, Shire D, Le Fur G, Casellas P (1995) Expression of central and peripheral cannabinoid receptors in human immune tissues and leukocyte subpopulations. Eur J Biochem **232**:54-61.
8. Bouaboula M, Rinaldi M, Carayon P, Carillon C, Delpech B, Shire D, Lefur G, Casellas P (1993) Cannabinoid-receptor expression in human leukocytes. Eur J Biochem **214**:173-180.
9. Kaminski NE, Abood ME, Kessler FK, Martin BR, Schatz AR (1992) Identification of a functionally relevant cannabinoid receptor on mouse spleen cells involved in cannabinoid-mediated immune modulation. Mol Pharmacol **42**:736-742.
10. Van Sickle MD, Duncan M, Kingsley PJ, Mouihate A, Urbani P, Mackie K, Stella N, Makriyannis A, Piomelli D, Davison JS, Marnett LJ, Di Marzo V, Pittman QJ, Patel KD,

Sharkey KA (2005) Identification and functional characterization of brainstem cannabinoid CB2 receptors. Science **310**:329-332.
11. Malan TP, Ibrahim MM, Deng HF, Liu Q, Mata HP, Vanderah T, Porreca F, Makriyannis A (2001) CB2 cannabinoid receptor-mediated peripheral antinociception. Pain **93**:239-245.
12. Guindon J, Hohmann A (2008) Cannabinoid CB2 receptors: A therapeutic target for the treatment of inflammatory and neuropathic pain. Br J Pharmacol **153**:319-334.
13. Beltramo M, Bernardini N, Bertorelli R, Campanella M, Nicolussi E, Fredduzzi S, Reggiani A (2006) CB2 receptor-mediated antihyperalgesia: Possible direct involvement of neural mechanisms. Eur J Neurosci **23**:1530-1538.
14. Bingham B, Jones PG, Uveges AJ, Kotnis S, Lu P, Smith VA, Sun SC, Resnick L, Chlenov M, He Y, Strassle BW, Cummons TA, Piesla MJ, Harrison JE, Whiteside GT, Kennedy JD (2007) Species-specific in vitro pharmacological effects of the cannabinoid receptor 2 (CB2) selective ligand AM1241 and its resolved enantiomers. Br J Pharmacol **151**:1061-1070.
15. Hohmann AG, Farthing JN, Zvonok AM, Makriyannis A (2004) Selective activation of cannabinoid CB2 receptors suppresses hyperalgesia evoked by intradermal capsaicin. J Pharmacol Exp Ther **308**:446-453.
16. Ibrahim MM, Deng HF, Zvonok A, Cockayne DA, Kwan J, Mata HP, Vanderah TW, Lai J, Porreca F, Makriyannis A, Malan TP (2003) Activation of CB2 cannabinoid receptors by AM1241 inhibits experimental neuropathic pain: Pain inhibition by receptors not present in the CNS. Proc Natl Acad Sci USA **100**:10529-10533.
17. Ibrahim MM, Porreca F, Lai J, Albrecht PJ, Rice FL, Khodorova A, Davar G, Makriyannis A, Vanderah TW, Mata HP, Malan TP (2005) CB2 cannabinoid receptor activation produces antinociception by stimulating peripheral release of endogenous opioids. Proc Natl Acad Sci USA **102**:3093-3098.
18. Ibrahim MM, Rude ML, Stagg NJ, Mata HP, Lai J, Vanderah TW, Porreca F, Buckley NE, Makriyannis A, Malan TP (2006) CB2 cannabinoid receptor mediation of antinociception. Pain **122**:36-42.
19. Nackley AG, Suplita RL, Hohmann AG (2003) A peripheral cannabinoid mechanism suppresses spinal Fos protein expression and pain behavior in a rat model of inflammation. Neuroscience **117**:659-670.
20. Nackley AG, Zvonok AM, Makriyannis A, Hohmann AG (2004) Activation of cannabinoid CB(2) receptors suppresses C-fiber responses and windup in spinal wide dynamic range neurons in the absence and presence of inflammation. J Neurophysiol **92**:3562-3574.
21. Quartilho A, Mata HP, Ibrahim MM, Vanderah TW, Porreca F, Makriyannis A, Malan TP (2003) Inhibition of inflammatory hyperalgesia by activation of peripheral CB2 cannabinoid receptors. Anesthesiology **99**:955-960.
22. Malan Jr. T, Ibrahim M, Lai J, Vanderah T, Makriyannis A, Porreca F (2003) CB cannabinoid receptor agonists: Pain relief without psychoactive effects? Curr Opin Pharm **3**:62-67.
23. Herzberg U, Eliav E, Bennett GJ, Kopin IJ (1997) The analgesic effects of R(+)-WIN 55,212-2 mesylate, a high affinity cannabinoid agonist, in a rat model of neuropathic pain. Neurosci Lett **221**:157-160.
24. Bridges D, Ahmad K, Rice ASC (2001) The synthetic cannabinoid WIN55,212-2 attenuates hyperalgesia and allodynia in a rat model of neuropathic pain. Br J Pharmacol **133**:586-594.
25. Pryce G, Baker D (2007) Control of spasticity in a multiple sclerosis model is mediated by CB(1), not CB(2), cannabinoid receptors. Br J Pharmacol **150**:519-525.
26. Choong KC, Su X, Urban MO (2007) Effect of CP55,940 on mechanosensory spinal neurons following chronic inflammation. Neurosci Lett **414**:105-109.
27. Hollinshead SP, Tidwell MW, Palmer J, Guidetti R, Sanderson A, Johnson MP, Chambers MG, Oskins J, Stratford R, Astles PC (2013) Selective cannabinoid receptor type 2 (CB2)

agonists: Optimization of a series of purines leading to the identification of a clinical candidate for the treatment of osteoarthritic pain. J Med Chem **56**:5722-5733.
28. Abrol R, Bray JK, Goddard WA, III (2012) BiHelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. Proteins: Struct Funct Bioinform **80**:505-518.
29. Abrol R, Griffith AR, Bray JK, Goddard WA, III. Structure prediction of G protein-coupled receptors and their ensemble of functionally important conformations. In: Vaidehi N, Klein-Seetharaman J, Eds. (2012) Membrane Protein Structure: Methods and Protocols. Humana Press, New York, NY.
30. Abrol R, Kim S-K, Bray JK, Griffith AR, Goddard WA, III (2011) Characterizing and predicting the functional and conformational diversity of seven-transmembrane proteins. Methods **55**:405-414.
31. Abrol R, Kim S-K, Bray JK, Trzaskowski B, Goddard WA, III (2013) Conformational ensemble view of G protein-coupled receptors and the effect of mutations and ligand binding. Methods Enzymol **520**:31-48.
32. Bray JK, Abrol R, Goddard WA, III, Trzaskowski B, Scott CE (2014) SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations. Proc Natl Acad Sci USA **111**:E72-E78.
33. Ahn KH, Scott CE, Abrol R, Goddard WA, III, Kendall DA (2013) Computationally-predicted CB1 cannabinoid receptor mutants show distinct patterns of salt-bridges that correlate with their level of constitutive activity reflected in G protein coupling levels, thermal stability, and ligand binding. Proteins: Struct Funct Bioinform **81**:1304-1317.
34. Scott CE, Abrol R, Ahn KH, Kendall DA, III WAG (2013) Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations. Protein Sci **22**:101-113.
35. Ballesteros JA, Weinstein H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In: Sealfon SC, Ed. (1995) Receptor Molecular Biology. Academic Press, Waltham, Massachusetts, pp. 366-428.
36. D'Antona A, Ahn KH, Kendall DA (2006) Mutations of CB_1 T210 produce active and inactive receptor forms: Correlations with ligand affinity, receptor stability, and cellular localization. Biochemistry **45**:5606-5617.
37. D'Antona A, Ahn KH, Wang L, Mierke DF, Lucas-Lenard J, Kendall DA (2006) A cannabinoid receptor 1 mutation proximal to the DRY motif results in constitutive activity and reveals intramolecular interactions involved in receptor activation. Brain Res **1108**:1-11.
38. Kam VWT, Goddard WA (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. Journal of Chemical Theory and Computation **4**:2160-2169.
39. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res **25**:3389-3402.
40. Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, Desale H, Clemons B, Cahalan SM, Schuerer SC, Sanna MG, Han GW, Kuhn P, Rosen H, Stevens RC (2012) Crystal structure of a lipid G protein-coupled receptor. Science **335**:851-855.
41. Chien EYT, Liu W, Zhao QA, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, Stevens RC (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. Science **330**:1091-1095.
42. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG, Schertler GFX (2008) Structure of a β_1-adrenergic G-protein-coupled receptor. Nature **454**:486-491.

43. Rasmussen SGF, Choi H-J, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VRP, Sanishvili R, Fischetti RF, Schertler GFX, Weis WI, Kobilka BK (2007) Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. Nature **450**:383-387.
44. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H-J, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human β_2-adrenergic G protein–coupled receptor. Science **318**:1258-1265.
45. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science **330**:1066-1071.
46. Okada T, Le Trong I, Fox BA, Behnke CA, Stenkamp RE, Palczewski K (2000) X-ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. J Struct Biol **130**:73-80.
47. Teller DC, Okada T, Behnke CA, Palczewski K, Stenkamp RE (2001) Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). Biochemistry **40**:7761-7772.
48. Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y (2002) Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. Proc Natl Acad Sci USA **99**:5982-5987.
49. Li J, Edwards PC, Burghammer M, Villa C, Schertler GFX (2004) Structure of bovine rhodopsin in a trigonal crystal form. J Mol Biol **343**:1409-1438.
50. Okada T, Sugihara M, Bondar A-N, Elstner M, Entel P, Buss V (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. J Mol Biol **342**:571-583.
51. Standfuss J, Xie G, Edwards PC, Burghammer M, Oprian DD, Schertler GFX (2007) Crystal structure of a thermally stable rhodopsin mutant. J Mol Biol **372**:1179-1188.
52. Stenkamp RE (2008) Alternative models for two crystal structures of bovine rhodopsin. Acta Crystallogr D **64**:902-904.
53. Murakami M, Kouyama T (2008) Crystal structure of squid rhodopsin. Nature **453**:363-367.
54. Shimamura T, Hiraki K, Takahashi N, Hori T, Ago H, Masuda K, Takio K, Ishiguro M, Miyano M (2008) Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. J Biol Chem **283**:17753-17756.
55. Doré AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, Hurrell E, Bennett K, Congreve M, Magnani F, Tate CG, Weir M, Marshall FH (2011) Structure of the adenosine A_{2A} receptor in complex with ZM241385 and the xanthines XAC and caffeine. Structure **19**:1283-1293.
56. Moukhametzianov R, Warne T, Edwards PC, Serrano-Vega MJ, Leslie AGW, Tate CG, Schertler GFX (2011) Two distinct conformations of helix 6 observed in antagonist-bound structures of a β_1-adrenergic receptor. Proc Natl Acad Sci USA **108**:8228-8232.
57. Yao XJ, Parnot C, Deupi X, Ratnala VRP, Swaminath G, Farrens D, Kobilka B (2006) Coupling ligand structure to specific conformational switches in the β_2-adrenoceptor. Nat Chem Biol **2**:417-422.
58. Ballesteros JA, Jensen AD, Liapakis G, Rasmussen SGF, Shi L, Gether U, Javitch JA (2001) Activation of the β_2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. J Biol Chem **276**:29171-29177.
59. Rhee MH, Nevo I, Levy R, Vogel Z (2000) Role of the highly conserved Asp-Arg-Tyr motif in signal transduction of the CB_2 cannabinoid receptor. FEBS Lett **466**:300-304.
60. Feng WK, Song ZH (2003) Effects of D3.49A, R3.50A, and A6.34E mutations on ligand binding and activation of the cannabinoid-2 (CB2) receptor. Biochem Pharmacol **65**:1077-1085.

61. Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman AP, Stevens RC (2008) The 2.6 Å crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. Science **322**:1211-1217.
62. Park JH, Scheerer P, Hofmann KP, Choe H-W, Ernst OP (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature **454**:183-187.
63. Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP, Ernst OP (2008) Crystal structure of opsin in its G-protein-interacting conformation. Nature **455**:497-502.
64. Jin WZ, Brown S, Roche JP, Hsieh C, Celver JP, Kovoor A, Chavkin C, Mackie K (1999) Distinct domains of the CB1 cannabinoid receptor mediate desensitization and internalization. J Neurosci **19**:3773-3780.
65. Choe HW, Kim YJ, Park JH, Morizumi T, Pai EF, Krauss N, Hofmann KP, Scheerer P, Ernst OP (2011) Crystal structure of metarhodopsin II. Nature **471**:651-655.
66. Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK (2011) Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. Nature **469**:175-180.
67. Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG (2011) Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. Nature **474**:521-525.
68. Xu F, Wu HX, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC (2011) Structure of an agonist-bound human A_{2A} adenosine receptor. Science **332**:322-327.
69. Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK (2011) Crystal structure of the β_2 adrenergic receptor-G_s protein complex. Nature **477**:549-555.
70. Kenakin T (2002) Drug efficacy at G protein-coupled receptors. Annu Rev Pharmacol Toxicol **42**:349-379.
71. Kenakin T (2002) Efficacy at G-protein-coupled receptors. Nat Rev Drug Discov **1**:103-110.
72. Kenakin T (2003) Ligand-selective receptor conformations revisited: The promise and the problem. Trends Pharmacol Sci **24**:346-354.
73. Kenakin T, Miller LJ (2010) Seven transmembrane receptors as shapeshifting proteins: The impact of allosteric modulation and functional selectivity on new drug discovery. Pharmacol Rev **62**:265-304.
74. Howlett AC, Fleming RM (1984) Cannabinoid inhibition of adenylate-cyclase. Pharmacology of the response in neuro-blastoma cell-membranes. Mol Pharmacol **26**:532-538.
75. Abadji V, Lucas-Lenard JM, Chin CN, Kendall DA (1999) Involvement of the carboxyl terminus of the third intracellular loop of the cannabinoid CB1 receptor in constitutive activation of G(s). J Neurochem **72**:2032-2038.
76. Glass M, Felder CC (1997) Concurrent stimulation of cannabinoid CB1 and dopamine D2 receptors augments cAMP accumulation in striatal neurons: Evidence for a G(s) linkage to the CB1 receptor. J Neurosci **17**:5327-5333.
77. Daigle TL, Kwok ML, Mackie K (2008) Regulation of CB(1) cannabinoid receptor internalization by a promiscuous phosphorylation-dependent mechanism. J Neurochem **106**:70-82.
78. Kouznetsova M, Kelley B, Shen MX, Thayer SA (2002) Desensitization of cannabinoid-mediated presynaptic inhibition of neurotransmission between rat hippocampal neurons in culture. Mol Pharmacol **61**:477-485.
79. Breivogel CS, Lambert JM, Gerfin S, Huffman JW, Razdan RK (2008) Sensitivity to Delta 9-tetrahydrocannabinol is selectively enhanced in beta-arrestin2-/- mice. Behav Pharmacol **19**:298-307.

80. Bonhaus DW, Chang LK, Kwan J, Martin GR (1998) Dual activation and inhibition of adenylyl cyclase by cannabinoid receptor agonists: Evidence for agonist-specific trafficking of intracellular responses. J Pharmacol Exp Ther **287**:884-888.

Chapter VII

STRUCTURE PREDICTION OF THE HUMAN MU-OPIOID RECEPTOR AND CONSTITUTIVELY ACTIVE MUTANTS

The majority of the human mu-opioid structure prediction computational experiments were performed by Nicholas Parker as part of the Caltech Freshman Summer Research Institute (FSRI).

ABSTRACT

The human mu-opioid receptor (hMOR) plays an important role in mediating pain and is the target for many drugs such as morphine and codeine. Designing drugs that interact with this receptor has proven difficult because its structure was not crystalized until 2012. Even then, the crystal structure captures a single pose of a dynamic molecule. Before the mouse mu-opioid receptor (mMOR) was crystallized, we used the Monte Carlo GEnSeMBLE method to predict the structure of the highly conserved hMOR using the crystallized human chemokine CXCR4 receptor as the template. Our lowest energy prediction closely resembled the crystallized mMOR since it had a backbone RMSD of 1.94 Å, well within the crystal resolution of 2.80 Å. The predicted hMOR also contained many of the same hydrogen bond networks observed in the crystallized GPCR. However, this mMOR was crystallized with a covalently bound antagonist, which may impact the structure of the receptor. Since our GEnSeMBLE method accurately produced the hMOR structure, we used it to perform a local sampling on the mMOR crystal structure and found five structures that are more energetically favorable than the crystal structure. Not surprisingly, the transmembrane region (TM) with the largest variations in rotation and sweep angles is TM5, which contains the residue covalently linked to the antagonist. These findings suggest that although the crystal structure is accurate, it may not be the best representation of the receptor without the antagonist present. Finally, we used the homologized mMOR as a template to build the hMOR wild-type (WT) receptor as well as five other mutants that have been experimentally determined to be functionally significant in the rat mu-opioid (rMOR) sequence. The predicted structures of the six receptors have different patterns of salt bridges and hydrogen bonds that can explain the functional differences observed in the experiments. For example, the highly constitutively active mutant T6.34K is the only receptor that has a hydrogen bond between a threonine on the intracellular end of TM2 and the aspartic acid of the highly conserved DRY motif. This polar connection may be important for stabilizing the active conformation. Thus, not only have we predicted the structure of hMOR but of its mutants that represent the constitutively active form of the receptor. By understanding the different structures at various functional levels, in the future, we can design drugs that induce or stabilize these different conformations.

INTRODUCTION

Morphine is one of the most effective pain medications available, but it is highly addictive and induces tolerance and dependency as well as having serious side effects such as respiratory depression (1). This drug primarily targets the mu-opioid receptor (MOR) as evidenced by MOR knock-out mice who do not experience the analgesic effects or physical dependence associated with morphine (2). Obviously, a pain medication with less serious side effects is desirable, but none has yet been developed or widely used. The MOR is a G protein-coupled receptor (GPCR) that has seven alpha-helical transmembrane regions (TMs) that span the cellular bilayer. This receptor primarily binds to the intracellular G_i protein, which inhibits the production of cyclic adenosine monophosphate (cAMP), but only one G_i protein subunit subtype, $G_i2\alpha$, is responsible for the antinociception effects caused by morphine (3). Perhaps it is possible to design a drug that evokes the signaling pathway for pain mediation, but not the ones associated with other undesirable side effects. To do this, the drug would have to stabilize a particular MOR conformation that binds to a specific intracellular signaling receptor. However such drug design has proven difficult because MOR was not crystallized until 2012, and until then, its structure was unknown.

That year, the four members of the opioid family were crystallized including the mouse MOR (mMOR) (4), the mouse delta-opioid receptor (5), the human kappa-opioid receptor (6), and the human nociception/orphanin FQ receptor (7). Despite the wealth of structural information, the crystal structures are just a single conformation of a very flexible and dynamic macromolecule that binds to a variety of intracellular proteins and different types of agonists (3; 8; 9). Thus, our lab has developed computational methodology to predict an ensemble, rather than a single conformation, of energetically accessible GPCR structures. The strength in this process is that these programs quickly and efficiently calculate the energies of a trillion possible structures by rotating and tilting the seven alpha-helices. Unlike homology methods, this procedure is sensitive enough to detect the structural impact of a single residue as proven by our previous applications to the human cannabinoid type 1 (CB1) receptor and its respective constitutively active mutants (CAMs) (10-12). We applied similar procedures to explain the CAMs of the rat MOR (rMOR) shown in **Table 7.1** (13). A single or double point mutation showed significant changes in the binding affinity for the DAMGO opioid agonist. In two cases, the L6.30E/T6.34K and T6.34K had a substantial increase in GTPγS assays compared to the wild-type (WT) rMOR receptor, which indicates an increase in basal level constitutive activity. These rMOR single point or

double point CAMs have not been crystallized, so it is unknown how these mutations affect the structures.

Table 7.1 Constitutively active mutants (CAMs) for the rMOR sequence (13).

rMOR Receptor	K_i Ratio Mutant:WT	Agonist DAMGO K_i (nM)	Basal GTPγS binding (% of WT)	Relative activity level
L6.30E/T6.34D	447:1	3189	100	Very inactive
T6.34D	96:1	684	100	Less active
L6.30E	93:1	667	100	Less active
WT	1:1	7.14	100	----------
L6.30E/T6.34K	1:1	6.98	124	Same as WT
T6.34K	1:21	0.34	210	More active

Here, we use the GEnSeMBLE (GPCR Ensemble of Structures in Membrane BiLayer Environment) programs (11; 14-17) to predict the structure of the human MOR (hMOR). The resulting conformations are similar to the crystallized structure of mMOR in that the lowest energy hMOR structure has a backbone RMSD of 1.94 Å. However, the mMOR was crystallized with the covalently bound β-funaltrexamine (β-FNA), which may affect the overall structure, so we used our methods to preform a local sampling of the crystal structure and found five conformations that are energetically more favorable than the crystallized one. Then using the mMOR helices as templates, the structures of the hMOR and the five mutants listed in **Table 7.1** are predicted. After a thorough sampling, the resulting receptors are very different, and these structural differences can explain the experimental differences in function. By understanding the different structures, in the future, drugs can be designed to interact with specific receptor conformations for a desired function.

METHODS

(For more details about structure prediction methodology, see *Chapter II*.)

hMOR structure prediction

The methods described herein are similar to the ones used for the CB1 (10; 12) and CB2 receptors discussed in *Chapters III* and *VI* respectively (**Figure 7.1**). We used the Monte Carlo GEnSeMBLE method (11; 14-17) developed within our lab to perform a thorough sampling of helical rotation (eta, η), tilt (theta, θ), and sweep (phi, φ) angles by quickly and efficiently calculating the energies of a trillion structures. The shape of the helices were constructed in two

ways—with OptHelix helices, or helices originally constructed from poly-alanine canonical helices and modified with molecular dynamics (MD) simulations and minimization; or with homologized helices, or helices from a crystallized GPCR mutated to match the target sequence.

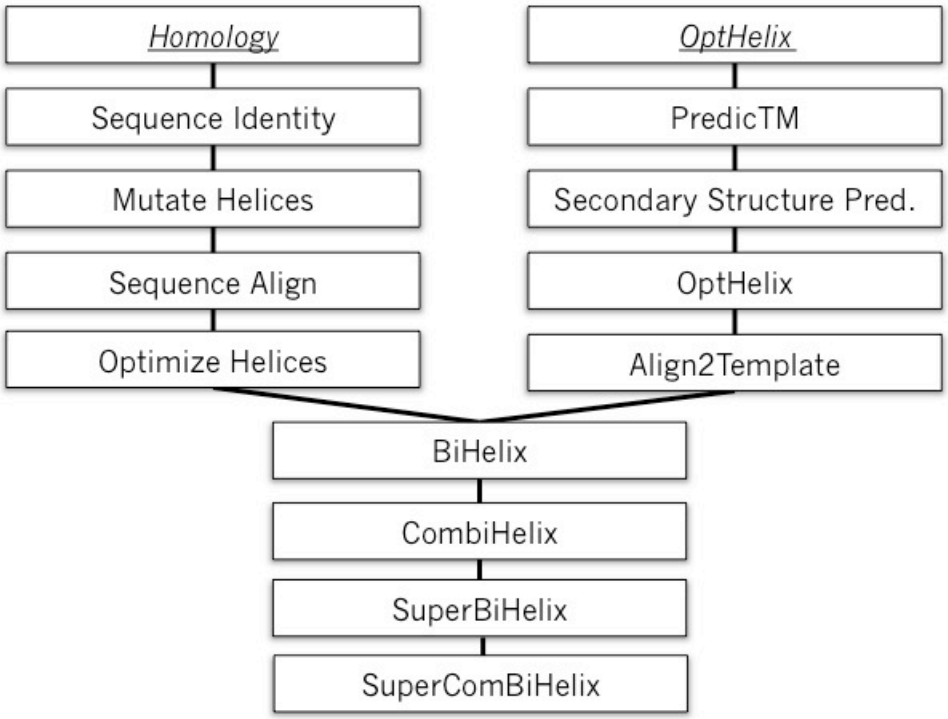


Figure 7.1 Schematic illustration explaining the computational GEnSeMBLE methodology used for predicting the hMOR structure. Figure courtesy of Nicholas Parker.

hMOR structure prediction with OptHelix helices

For the OptHelix helices, we had to first determine the respective lengths of the seven individual helices. A BLAST search (18) with an E-value of 0.1 found 1834 sequences related to hMOR ranging in TM sequence identity from 100% to 7.75%. After using MAFFT software (19) to align these sequences, a hydrophobicity value based on the Wimley-White scale (20) was assigned to each residue, and these values were averaged over all 1834 sequences to get an average hydrophobic value for every residue in the hMOR sequence. The program PredicTM identified seven hydrophobic regions that would correspond to the seven TM regions spanning the greasy bilayer. However, crystallized GPCRs have shown that the helical portions of the TMs extend beyond the lipid bilayer into the solvent. Secondary structure prediction (SSP) servers PORTER (21; 22), APSSP2 (23; 24), and PSIPRED (25; 26) were used to predict the helical

regions and extend the hydrophobic portions identified with PredicTM. Once the TM lengths have been determined, the helices were built using OptHelix. OptHelix constructed the TMs initially with canonical helices made entirely of alanine residues. The alanines were replaced with the rotamers of the correct residues over the course of the program using the Side chain Rotamer Excitation Analysis Method (SCREAM) (27). The individual helices were stabilized with MD, and two structures were collected based on the lowest energy structure of the simulation (MinEng) and the structure with whose root-mean-square deviation (RMSD) is closest to the average structure of the simulation (MinRMSD). Once the seven helices were built, they were aligned to one of the crystallized GPCRs listed in **Table 7.2** or the predicted structure of the human chemokine CCR5 receptor (hCCR5). There were two ways to align the TMs to the templates. The first way was to align the residue representing the midpoint of the raw hydrophobic region of the TM to the midplane bisecting the TM bundle (rawmid). The second way was to align the residue bisecting the area of the hydrophobic region of the TM to the bundle midplane (area).

Table 7.2 Comparison of the TM sequence identity of hMOR with crystallized GPCR sequences including hCXCR4 (28), bRho (29), Hβ_2AR (30), HA$_{2A}$AR (31), and hCCR5 (32-34).

Crystallized Receptor	Percent TM sequence identity with respect to hMOR
Human Mu-Opioid Receptor	100%
Human Chemokine CXCR4 Receptor	46.96%
Bovine Rhodopsin Receptor	45.56%
Human β_2 Adrenergic Receptor	37.79%
Human Adenosine A2A Receptor	34.68%
Human Chemokine CCR5 Receptor	26.89%

hMOR structure prediction with homologized helices

Rather than constructing individual helices from scratch, we also mutated crystallized helices so that they matched the target sequence and then optimized the helices. Unlike homologized helices, OptHelix helices were constructed individually, so their respective shapes were not affected by interactions with the other TMs or the loop regions. Since the receptor sequence affects the helical shape, the homologized helices produce favorable energies for target receptors that they are closely related to. Fortunately, when this project began, there were already multiple crystallized GPCRs with a high percentage of conserved sequence identity in the TM regions (**Table 7.2**). A high sequence identity is anything greater than 30%, and there are two structures with TM sequence identities at 45% or greater. Thus, there are four crystallized GPCRs that were appropriate homology templates including the human chemokine CXCR4 receptor (hCXCR4)

(28), the bovine rhodopsin receptor (bRho) (29), the human β_2 adrenergic receptor ($H\beta_2AR$) (35), and the human A_{2A} adenosine receptor ($HA_{2A}AR$) (31). The GEnSeMBLE predicted structure of hCCR5 was also included despite its low sequence identity and the fact it had not been crystallized at the time. It was included because it is closely related to hCXCR4, which had the highest sequence identity with the target sequence. Also, even though hCCR5 had not been crystallized, it agreed with a sufficient amount of experimental evidence (32-34). In 2013, hCCR5 was crystallized (36), and it was in good agreement with the predicted structure.

Sampling the helical orientations

After constructing the helical bundles, we performed a complete sampling using the BiHelix and ComBiHelix programs (14). In BiHelix, pairs of interacting helices were rotated 360° in 30° increments. The pairwise energies were calculated and summed up to obtain an estimate of the energies of the entire bundle. Then, the 2000 lowest energy estimates were constructed as complete 7TM bundles in ComBiHelix, and the charged interhelical and the charged total energies were calculated. The interhelical energy is equal to the difference in energy of the entire bundle and the sum of the energies of the seven individual helices. **Table 7.3** shows the lowest energy conformations for each template. There are significant changes in helical orientations for all the templates except $H\beta_2AR$ with homologized helices and hCXCR4 with homologized helices, and slight changes in $HA_{2A}AR$ with homologized helices and hCCR5 with homologized helices. The hMOR structure with hCXCR4 homologized helices had the most favorable average of the charged and neutralized interhelical energies by over 100 kcal/mol. Thus, we used the top three structures from the hCXCR4 homologized helices ComBiHelix results according to average energy rank for further local sampling in the SuperBiHelix and SuperComBiHelix programs (11). The theta (θ, tilt) angles were sampled in from -10° to 10° in 10° increments and the phi (φ, sweep) and the eta (η, rotation) angles were sampled from -30° to +30° in 15° increments. If the last two residues of the helix were charged, they were replaced with alanine residues using the SCREAM program (27). The 2000 lowest energy estimates for each starting hCXCR4 structure were constructed with the correct sequence in SuperComBiHelix, and the charged total energy, the neutralized total energy, the charged interhelical energy, and the neutralized interhelical energy were calculated. The resulting structures were ranked according to these four criteria, and they were averaged to give a final ranking. The structures of the three starting points were combined and re-ranked. The ten lowest average energy rank structures were the result of sampling the lowest energy ComBiHelix structure.

Table 7.3 Interhelical energy comparison for hMOR with OptHelix helices and homologized helices. Optimized eta (η, rotation) angles from the ComBiHelix analysis for the WT hMOR receptors with various helices: OptHelix helices aligned to the crystallized Hβ_2AR (PDB ID: 2RH1) (30), the crystallized bRho (PDB ID: 1U19) (29), the crystalized hCXCR4 (PDB ID: 3OE6) (28), the crystallized HA_{2A}AR (PDB ID: 3EML) (31), and the predicted hCCR5 structure (32-34), as well as homologized helices with the same templates.

Helix Method	Template	Helix Type	Hydrophobic Center	TM1 Eta	TM2 Eta	TM3 Eta	TM4 Eta	TM5 Eta	TM6 Eta	TM7 Eta	Charge Interhelical Energy	Neutral Interhelical Energy	Average Interhelical Energy
Homology	Hβ2AR	Homology	Homology	0	0	0	0	0	0	0	-345.6	-391.9	-368.8
OptHelix	Hβ2AR	MinEng	Area	0	-60	150	30	-120	-120	90	-306.6	-296.3	-301.5
OptHelix	Hβ2AR	MinEng	Rawmid	0	-60	150	30	-150	60	120	-268.1	-288.0	-278.1
OptHelix	Hβ2AR	MinRMSD	Area	0	-90	-120	0	-120	-120	60	-304.7	-309.6	-307.2
OptHelix	Hβ2AR	MinRMSD	Rawmid	30	-90	150	0	-150	120	180	-287.5	-295.4	-291.5
Homology	bRhod	Homology	Homology	0	0	0	30	180	-30	0	-350.8	-316.2	-333.5
OptHelix	bRhod	MinEng	Area	-60	-150	90	120	-150	120	-120	-209.1	-253.1	-231.1
OptHelix	bRhod	MinEng	Rawmid	-30	180	120	30	-150	-120	-60	-263.0	-244.0	-253.5
OptHelix	bRhod	MinRMSD	Area	30	-150	150	0	180	120	-120	-257.3	-265.0	-261.2
OptHelix	bRhod	MinRMSD	Rawmid	0	-90	150	30	120	120	-120	-276.6	-273.4	-275.0
Homology	hCXCR4	Homology	Homology	0	0	0	0	0	0	0	-523.1	-430.3	-476.7
OptHelix	hCXCR4	MinEng	Area	0	150	-60	60	180	90	-150	-288.5	-266.0	-277.3
OptHelix	hCXCR4	MinEng	Rawmid	-150	150	60	60	30	150	-90	-80.4	-429.1	-254.8
OptHelix	hCXCR4	MinRMSD	Area	0	180	-150	30	-30	-90	-60	-243.3	-277.6	-260.5
OptHelix	hCXCR4	MinRMSD	Rawmid	-30	180	-60	-30	60	-120	-60	-288.7	-291.5	-290.1
Homology	hAA2A	Homology	Homology	0	0	0	0	-90	0	0	-359.3	-366.2	-362.7
OptHelix	hAA2A	MinEng	Area	30	60	-60	60	-60	120	90	-239.7	-206.2	-223.0
OptHelix	hAA2A	MinEng	Rawmid	-60	-120	30	0	-30	30	90	-254.7	-226.6	-240.7
OptHelix	hAA2A	MinRMSD	Area	30	30	180	-30	-30	120	60	-234.1	-207.9	-221.0
OptHelix	hAA2A	MinRMSD	Rawmid	-60	-120	-30	60	-60	120	90	-270.5	-234.1	-252.3
Homology	CCR5	Homology	Homology	-30	30	0	60	0	0	0	-323.1	-376.7	-349.9
OptHelix	CCR5	MinEng	Area	60	-60	0	30	-90	-90	0	-285.6	-309.9	-297.8
OptHelix	CCR5	MinEng	Rawmid	-60	120	90	0	-120	-150	-90	-301.8	-288.7	-295.3
OptHelix	CCR5	MinRMSD	Area	-30	-120	90	-30	-150	120	120	-293.7	-303.1	-298.4

Local sampling of mMOR crystal structure

To test the flexibility of the mMOR (PDB ID: 4DKL) (4) without the antagonist present, we preformed a local sampling on the crystallized structure using the angles specified in **Table 7.4**.

The phi (φ, sweep) and eta (η, rotation) angle sampling range was wider than usual because the range was extended until none of the low energy structures had angles that reached the end of the span, meaning we had exhausted sampling the possible helical orientations. During the SuperBiHelix stage, if the last two residues of the helix were charged, they were replaced with alanine residues. The 2000 structures with the lowest energy estimates in SuperBiHelix were constructed with the correct sequence in SuperComBiHelix and ranked according to average energy rank.

Table 7.4 Theta (θ, tilt), phi (φ, sweep), and eta (η, rotation) angles sampled during local sampling of mMOR crystal structure with SuperBiHelix and SuperComBiHelix programs.

TM	Theta angle sampling	Phi angle sampling	Eta angle sampling
1	-10, 0, 10	-30, -15, 0, 15, 30	-30, -15, 0, 15, 30
2	-10, 0, 10	-30, -15, 0, 15, 30	-30, -15, 0, 15, 30
3	-10, 0, 10	-30, -15, 0, 15, 30	-30, -15, 0, 15, 30
4	-10, 0, 10	-30, -15, 0, 15, 30	-30, -15, 0, 15, 30
5	-10, 0, 10	-45, -30, -15, 0, 15, 30	-30, -15, 0, 15, 30, 45, 60
6	-10, 0, 10	-30 ,-15, 0, 15, 30	-30, -15, 0, 15, 30
7	-10, 0, 10	-30, -15, 0, 15, 30	-30, -15, 0, 15, 30

hMOR CAMs structure prediction

Using the mMOR crystal structure, homologized helices were constructed for the WT hMOR as well as the five mutants listed in **Table 7.1**. The mutant residue rotamer, either a glutamic acid at position 6.30, an aspartic acid at 6.34, a lysine at 6.34, or a combination [using Ballesteros-Weinstein numbering (37)], was added with the SCREAM program. First, we performed a 360° helical sampling using the BiHelix and ComBiHelix programs. The 2000 bundles with the lowest energy estimates in BiHelix were constructed in their entirety in ComBiHelix and ranked according to their average energy ranks. The last two charged residues in each helix of the lowest average energy rank structure were replaced with alanine residues for the SuperBiHelix step. Again, the 2000 lowest energy estimates from SuperBiHelix were constructed in their entirety including the terminal charged residues for SuperComBiHelix. Generally, theta (θ, tilt) angles were sampled in from -10° to 10° in 10° increments and the phi (φ, sweep) and the eta (η, rotation) angles were sampled from -30° to +30° in 15° increments. If a structure sampled one of the angles at the end of the range, then another round of sampling was performed with an extended range. All of the SuperComBiHelix structures were combined and ranked according to their average energy rank to determine the lowest energy structures.

RESULTS AND DISCUSSION

hMOR structure prediction and comparison to crystallized mMOR

Shortly after predicting the structure of hMOR, the crystal structure of the mouse mMOR was published (PDB ID: 4DKL) (4). The mMOR has a 99.21% conserved TM sequence identity with respect to hMOR, so it is highly likely that the mMOR has an extremely similar structure. Our predicted lowest energy structure for the hMOR has a 1.94 Å backbone RMSD with respect to mMOR (**Table 7.5** and **Figure 7.2**). Of the lowest energy predicted structures, the conformation

ranked eighth according to lowest average energy rank (WT8) has the lowest backbone RMSD with 1.78 Å. The range in backbone RMSD in our top ten predicted structures according to average energy rank is 1.78 - 2.04 Å, which is a close range and smaller than 2.80 Å, the crystal resolution of the mMOR. Because of the high percentage of conserved sequence identity and the small backbone RMSD value with respect to the crystallized mMOR, we trust our predicted structures of hMOR. The structure with '0' for all the angles, the hCXCR4 homology model, was close to having the lowest RMSD at 1.80 Å, but it was ranked tenth in energy. Since the helical sampling produced a structure with a slightly smaller RMSD and more favorable energies, the GEnSeMBLE method improved upon the homology model and shows the importance of helical sampling.

Table 7.5 Changes in theta (θ, tilt) angles, phi (φ, sweep) angles, eta (η, rotation) angles, and backbone RMSD of low energy predicted hMOR structures with respect to the crystallized mMOR (PDB ID: 4DKL) (4).

Theta	TM1	TM2	TM3	TM4	TM5	TM6	TM7	Phi	TM1	TM2	TM3	TM4	TM5	TM6	TM7	Eta	TM1	TM2	TM3	TM4	TM5	TM6	TM7	7TM RMSD (Å)
Theta	0	0	0	0	0	0	0	Phi	0	0	0	15	-15	0	-15	Eta	0	0	0	15	15	0	0	1.94
Theta	0	0	0	0	0	0	0	Phi	0	0	0	0	-15	0	-15	Eta	0	0	0	0	15	0	0	1.98
Theta	0	0	0	0	0	0	0	Phi	0	0	0	0	-15	15	-15	Eta	0	0	0	0	15	0	0	2.02
Theta	0	0	0	0	0	0	0	Phi	0	0	0	15	-15	0	-15	Eta	0	0	-15	0	15	0	0	1.98
Theta	0	0	0	0	0	0	0	Phi	0	0	0	15	-15	0	-15	Eta	0	0	0	0	15	0	0	1.96
Theta	0	0	0	0	0	0	0	Phi	0	0	0	0	-15	0	0	Eta	0	0	0	0	15	0	0	1.90
Theta	0	0	0	0	0	0	0	Phi	0	0	0	30	-15	0	15	Eta	0	0	0	15	15	0	0	1.81
Theta	0	0	0	0	0	0	0	Phi	0	0	0	15	0	0	0	Eta	0	0	0	0	-15	0	0	1.78
Theta	0	0	0	0	0	0	0	Phi	0	0	0	-15	-15	0	-15	Eta	0	0	-15	0	15	0	0	2.04
Theta	0	0	0	0	0	0	0	Phi	0	0	0	0	0	0	0	Eta	0	0	0	0	0	0	0	1.80

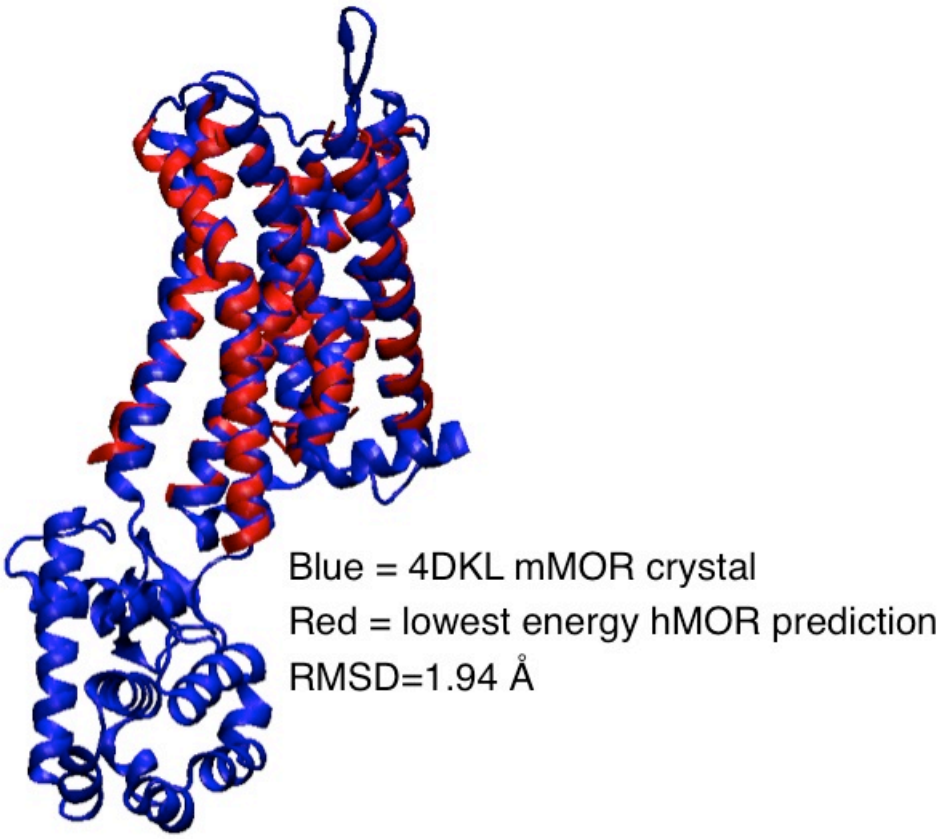


Figure 7.2 Comparison of the crystallized mMOR receptor (blue) (PDB ID: 4DKL) and the lowest energy predicted hMOR structure (red). Backbone RMSD of the TMs is 1.94 Å.

All ten predicted structures in **Table 7.5** have a TM1-TM2-TM3-TM7 hydrogen bond network similar to the ones shown for hMOR and the crystallized mMOR in **Figure 7.3**. Our hMOR structures predict that N1.50 forms a hydrogen bond with the S7.46 backbone amide oxygen, which is also observed in mMOR. In this crystal structure, there is a water-mediated hydrogen bond between D2.50 and N3.35. The predicted hMOR structure lacks water molecules, so instead there is a direct hydrogen bond between D2.50 and N3.35 in eight of the top ten predictions including the lowest energy and lowest RMSD poses. The crystal structure also shows that D2.50 has water-mediated interactions with N7.49 and Y7.53. In the absence of water, our lowest energy pose has a hydrogen bond between D2.50 and S7.46. The D2.50 residue is 5.94 Å away from N7.49—close enough for a water-mediated hydrogen bond. In most class A GPCRs, the N7.49 residue usually interacts with D2.50, but that is not the case in the mMOR crystal or the hMOR predicted structures. In the mMOR crystal, N7.39 is interacting with W6.48 via a water molecule, whereas in hMOR, there is a direct hydrogen bond. There is also a D3.32-

Y7.43 hydrogen bond in the predicted hMOR structures and the mMOR crystal structure as well as other crystallized GPCRs such as the Hβ_2AR (PDB ID: 2RH1) (30), the turkey β_1 adrenergic receptor (Tβ_1AR) (PDB ID: 2VT4) (38), the human dopamine D3 receptor (hDD3R) (PDB ID: 3PBL) (39), the human muscarinic M2 receptor (PDB ID: 3UON) (40), and the human muscarinic M3 receptor (PDB ID: 4DAJ) (41). (For the remainder of the chapter, hydrogen bonds will be indicated by a '-', and salt bridges will be indicated by a '+.')

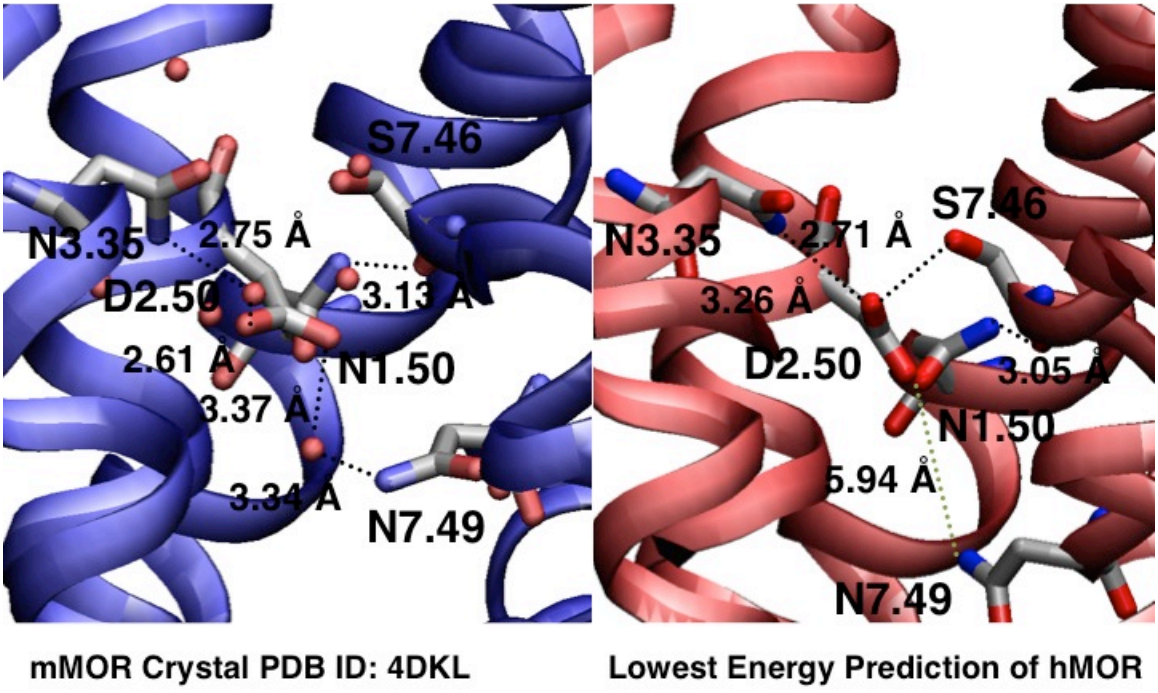


Figure 7.3 Comparison of the 1-2-3-7 hydrogen bond networks for the crystallized mMOR receptor (blue, left) and the lowest energy predicted structure of hMOR (red, right). Hydrogen bonds are indicated with black dotted lines, and their distances and interacting residues are labeled. The green dotted line on the right represents a possible water-mediated hydrogen bond. Water molecules are shown in the crystal structure on the left.

As for the TM2-TM3-TM4 hydrogen network, the mMOR crystal structure shows a very standard one of N2.45-W4.50 and N2.45-T3.42 connections. In the lowest energy hMOR predicted structure, there is a similar hydrogen bond between N2.45 and N4.46, which is one turn below the conserved tryptophan W4.50. The mMOR crystal structure has also a T3.42-N4.49 hydrogen bond, which is in the lowest RMSD predicted hMOR structure. In the lowest energy predicted hMOR structure, T3.42 is 3.34 Å away from N4.49, which is close enough for a possible hydrogen bond. In this structure, W4.50 has a different rotamer than in the mMOR crystal structure. The one in the lowest energy hMOR structure is perpendicular to the membrane rather

than parallel to it as in the crystal mMOR and lowest RMSD hMOR structures. However, in the mMOR crystal structure this W4.50 rotamer is forced into position by the 1-monooleoyl-rac-gylcerol molecule, which is absent from the predicted hMOR structures.

While the predicted hMOR structure and crystallized mMOR structure do have a lot of similar polar contacts and structural motifs, there are some differences between the two poses. Both of the lowest energy and lowest RMSD predicted hMOR structures show two salt bridge networks absent from the mMOR crystal (**Figure 7.4**). The first set involves the DRY motif on the intracellular end of TM3 (**Figure 7.4A**). In our structures we predict that D3.49 forms a salt bridge with R4.40 and a hydrogen bond with Y2.42. The salt bridge between D3.49 and R4.40 is similar to the one observed between D3.49 and R4.37 in hCXCR4 (28), D3.49 and R4.41 in the hDD3R (although the side chain of R4.41 was not resolved, PDB ID: 3PBL, chain B) (39), D3.49 and K4.41 in the predicted structure of CB1 (10; 12), and D3.49 and R4.41 in the predicted structure of CB2 (see *Chapter VI*). However, the presence of IC2 in the crystallized mMOR causes the intracellular portion of TM4 to wind up tighter than anticipated. This D3.49+R4.40 salt bridge does not form in the crystal, but instead D3.49 interacts with R179 of the second intracellular loop (IC2). The second set of salt bridges present in hMOR, but not mMOR is between K5.39+E5.35+K6.58 on the extracellular end (**Figure 7.4B**). In the mMOR crystal, K5.39 forms a covalent bond with the β-FNA ligand, so the K5.39 residue cannot form any salt bridge. Meanwhile, the K6.58 residue is forming a hydrogen bond with Q7.31, but the β-FNA ligand may be blocking the K6.58 residue's potential interaction with E5.35.

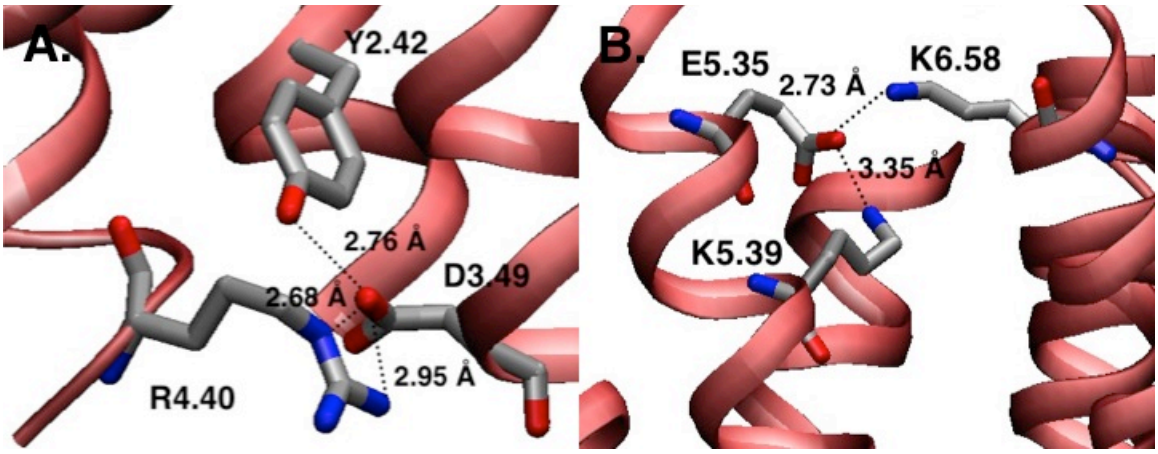


Figure 7.4 Hydrogen bond and salt bridge networks for **(A)** TMs 2, 3, and 4 and **(B)** TMs 5 and 6 for the lowest energy predicted structure of hMOR. Hydrogen bonds are indicated with black dotted lines, and their distances and interacting residues are labeled.

The crystallized mMOR contains a hydrogen bond not predicted in the hMOR structure. The mMOR structure has a hydrogen bond between R3.50 and T6.34. Since both the hMOR and mMOR lack the D6.30 conserved residue, this salt bridge between TM3 and TM6 may compensate for the lost ionic lock, which is believed to be important for preventing activation and is present in many crystallized GPCRs including bRho (29; 42-47), squid rhodopsin (48; 49), hDD3R (39), and thermally stabilized $HA_{2A}AR$ (50) and inactive $T\beta_1AR$ (51) and has been supported by experimental studies (52; 53). Perhaps this R3.50-T6.34 hydrogen bond is a polar contact important for stabilizing the inactive conformation, and is unique to the MORs.

The differences between the two structures described above, however, are very few, and the overall conformation of the receptors is very similar. Furthermore, the hMOR structure was predicted without loops, water molecules, or the β-FNA antagonist, which impact the residues' interactions and the TMs' backbone conformations. Even though there is a very high conserved sequence identity percentage between mMOR and hMOR, that does not guarantee that the two species' receptor structures are identical. As discussed below, a single mutation can have significant impact on the receptor's function and thereby structure (13). Yet, it is reasonable to expect a strong similarity between the two receptors' structures, which is what is observed and thereby gives credence to the hMOR predictions.

Local sampling of mMOR crystal structure: Identifying lower energy structures

As stated above, mMOR was crystallized in the presence of an antagonist, β-FNA, which is covalently bound to K5.39. It is assumed that the crystallization process captures the physiologically relevant receptor structure. According to Boltzmann's distribution, the lowest energy state is the most populated (54; 55). However, there are multiple issues to consider when analyzing crystal structures. First, GPCRs can have multiple inactive and active conformations (56), and crystal structures are just snapshots of a very dynamic and flexible macromolecule. Second, crystal packing can have an effect on the receptor structure, especially if receptors from different unit cells interact with one another. Third, multiple GPCRs, including mMOR, have been crystallized with a T4 lysozyme to replace IC3 and stabilize the receptor conformation. Yet, the fusion of the lysozyme protein to the GPCR may impact the structure, especially the attached TM5 and TM6. Fourth, ligands can impact the receptor structure so that it will not resemble the apo-form lacking a ligand. This last issue is rarely a problem, but it is particularly important for the β-FNA antagonist in mMOR. The covalent bond between the ligand and the receptor might impact the receptor structure, especially TM5. Here, we used SuperBiHelix to sample local theta

(θ, tilt), phi (φ, sweep) and eta (η, rotation) angles to obtain a more energetically favorable structure than the crystal conformation.

Table 7.6 shows the top ten average energy rank mMOR structures using homologized mMOR helices and extensive local sampling. The range of backbone RMSD is between 0.00 and 2.06 Å. None of the helices preferred a different tilt (θ) angle (data not shown in the table), but changes in the sweep (φ) angles ranged from -45º to 0º, and changes in the rotation (η) angles ranged from -15º to 45º. The vast majority of the helix sweeps and rotations occurred in TM5 as expected because β-FNA is not present to anchor or restrain TM5. TM5 rotates and/or sweeps as much as 45º, whereas TMs 4 and 6 sweep 15º, and TM3 rotates 15º. TMs 1 and 2 do not experience any tilt, sweep, or rotation. The crystal structure, represented by all '0º' for angles, is ranked sixth, which means five structures with more favorable energies were identified with local sampling. The backbone RMSD values of these five conformations is rather small with the top four having values less than 1.0 Å. The ten structures shown below have backbone RMSDs well below the crystal resolution of mMOR with 2.8 Å (4). However, even though the differences in the structures are slight, they are other more energetically favorable TM bundles identified via thorough sampling. Hence, the crystallized structure may not be the most accurate representation of mMOR.

Table 7.6 Predicted structures of mMOR after local sampling the crystallized mMOR (PDB ID: 4DKL) (4) structure with SuperBiHelix (11). Changes in phi, (φ, sweep) angles, eta (η, rotation) angles, and backbone RMSD of predicted mMOR 7TM bundles with respect to the crystallized mMOR (PDB ID: 4DKL) (4) are shown.

	φ (°)							η (°)							7TM RMSD
Rank	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM1	TM2	TM3	TM4	TM5	TM6	TM7	(Å)
1	0	0	0	0	0	-15	0	0	0	0	0	15	0	0	0.52
2	0	0	0	0	-15	0	0	0	0	0	0	0	0	0	0.69
3	0	0	0	-15	-15	-15	0	0	0	0	0	30	0	0	0.92
4	0	0	0	0	-15	0	0	0	0	-15	0	0	0	0	0.75
5	0	0	0	0	-30	0	0	0	0	0	0	45	0	0	1.48
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	-15	-15	0	0	0	0	0	30	0	0	0.91
8	0	0	0	0	-45	0	0	0	0	-15	0	45	0	0	2.06
9	0	0	0	-15	-30	0	0	0	0	0	0	45	0	0	1.49
10	0	0	0	-15	-45	0	0	0	0	0	0	45	0	0	2.05

Structural differences in hMOR CAMs can explain functional differences

Previous studies with CB1 have shown that our GEnSeMBLE structural prediction methods are sensitive enough to detect the conformational impact, which explains the functional impact, caused by a single point mutation (10; 12). Here, we apply these techniques to single and double point mutations to observe conformational changes that could possibly explain changes in agonist binding affinity (**Table 7.1**). Furthermore, the experimental data that we are using for comparison was done with the rMOR sequence (13), but these residues are conserved in the hMOR sequence. Even though rMOR and hMOR share a 99.21% sequence identity in the TMs, a single point mutation can have a serious affect on function as the experiment discussed here proves, so the goal of the analysis below is to predict if the trends in the rat sequence experiments would be applicable to the human one.

Since the crystallized mMOR receptor was available, we used it to create homologized helices for the WT hMOR and its five single and double point mutants. At the ComBiHelix stage, which consists of sampling 360º of rotation angles for the entire 7TM bundle, the majority of the most energetically favorable receptors' helices match the crystal structure (**Table 7.7**). The only two receptors that deviated from the crystal structure were the double mutants, L6.30E/T6.34D and L6.30E/T6.34K. Not surprisingly, the only TM that changed was TM5, and the receptor that had the largest RMSD with respect to the crystal structure was L6.30E/T6.34K, which was experimentally shown to have a significant binding affinity for the peptide agonist DAMGO and for GTPγS, and thus be constitutively active. What is surprising is that this double mutant has the same affinity for the agonist DAMGO that the WT does, but it has a backbone RMSD of 1.84 Å with respect to the WT. It was expected that the agonist would have the same binding affinity for similar structures, but it could be that the agonist does not interact with TM5, which is the only difference between the two receptors. The L6.30E/T6.34D sees only a slight change of 30º in TM5, but experimentally, it experiences a significant decrease of agonist binding affinity of 447-fold with respect to the WT receptor. However, further structural refinement was completed with the SuperBiHelix and SuperComBiHelix programs.

Table 7.7 Predicted structures of hMOR and respective mutants with the best average energy rank at the ComBiHelix level using homologized helices from the crystallized mMOR (PDB ID: 4DKL) (4) template. Eta, η (rotation) angles, backbone RMSD of predicted mMOR 7TM bundles with respect to the WT hMOR are shown.

hMOR receptor	Relative activity level (13)	η (°)							7TM RMSD
		TM1	TM2	TM3	TM4	TM5	TM6	TM7	(Å)
L6.30E/ T6.34D	Very inactive	0	0	0	0	-30	0	0	0.54
T6.34D	Less active	0	0	0	0	0	0	0	0.00
L6.30E	Less active	0	0	0	0	0	0	0	0.00
WT	---------	0	0	0	0	0	0	0	0.00
L6.30E/ T6.34K	Same as WT	0	0	0	0	-150	0	0	1.84
T6.34K	More active	0	0	0	0	0	0	0	0.00

After sampling the tilt, sweep, and rotation angles, we developed a consensus of the hydrogen bonds and salt bridge patterns for the top 20 structures according to average energy rank for each of the six receptors. Those polar contacts involving the DRY motif at the intracellular end of TM3 are shown below in **Table 7.8**. These residues interact with those on TM6 to stabilize the inactive form and prevent activation and G protein binding. There are different intracellular polar contacts depending upon the mutation, which affects the structure and thus activity level of the receptor. We can explain the different activity levels based on the different interhelical contacts. The number of salt bridges and hydrogen bonds involving the DRY motif increases with decreasing activation levels. We predict that the L6.30E/T6.34D mutant has two salt bridges, R3.50+T6.34D and R3.50+L6.30E, and a hydrogen bond Y2.43-D3.49. The TM3+TM6 salt bridges include the ionic lock observed in other class A GPCRs and are effective at preventing activation. Four out of the six receptors have an intracellular TM3+TM6 salt bridge with the exception being the highly constitutively active T6.34K mutant and the constitutively active WT receptor, which does have an R3.50-T6.34 hydrogen bond in the crystallized mMOR. The Y2.43-D3.49 salt bridge is observed in all the receptors except the most active one, T6.34K. This receptor has a T2.39-R3.50 hydrogen bond, which is not observed in any of the other receptors. Perhaps this hydrogen bond is necessary for stabilizing the highly constitutively active form of hMOR. Even though these functional experiments were done with the rMOR, our structural predictions suggest that the hMOR would show similar trends in activity levels.

Table 7.8 Hydrogen bond and salt bridge patterns of the predicted structures of hMOR and respective mutants at the using homologized helices from the crystallized mMOR (PDB ID: 4DKL) (4) template.

hMOR receptor	Relative activity level (13)	Hydrogen Bond/Salt bridge Patterns involving DRY motif	# Salt bridges	# H-bonds
L6.30E/ T6.34D	**Very inactive**	R3.50+L6.30E, R3.50+T6.34D, Y2.42-D3.49	2	1
T6.34D	**Less active**	R3.50+T6.34D, Y2.42-D3.49	1	1
L6.30E	**Less active**	R3.50+L6.30E (ionic lock), R3.50-T6.43, Y2.42-D3.49	1	2
WT	**----------**	[R3.50-T6.43 (seen in crystal 4DKL, not predicted structures)], Y2.42-D3.49	0	1
L6.30E/ T6.34K	**Same activity as WT**	R3.50+L6.30E (ionic lock), Y2.42-D3.49	1	1
T6.34K	**More active**	T2.39-R3.50 (no 3+6 interaction)	0	1

Surprisingly, the WT and L6.30E/T6.34K receptors have the same level of GTPγS basal and agonist DAMGO binding affinity, but the predicted structures are very different. **Figure 7.5A** shows a comparison of the TM bundles with the WT receptor in red and the double mutant in green. TM5 tilts and rotates significantly, while TM6 experiences some tilting resulting in a backbone RMSD of 2.20 Å between the two structures. It is expected that two receptors with such similar functional behaviors would have similar structures as well, but straight homology modeling of the L6.30E/T6.34K sequence in the WT structure (**Figure 7.5A**) shows that the resulting conformation is not energetically favorable. In the straight homology model, T6.34 is replaced with the longer side chain of lysine, which would block the formation of the ionic lock between R3.50 and L6.30E. Also, the T6.34K side chain would clash with M5.61. To alleviate high energies resulting from residue clashes, TM5 tilts and rotates to provide space for the long lysine side chain.

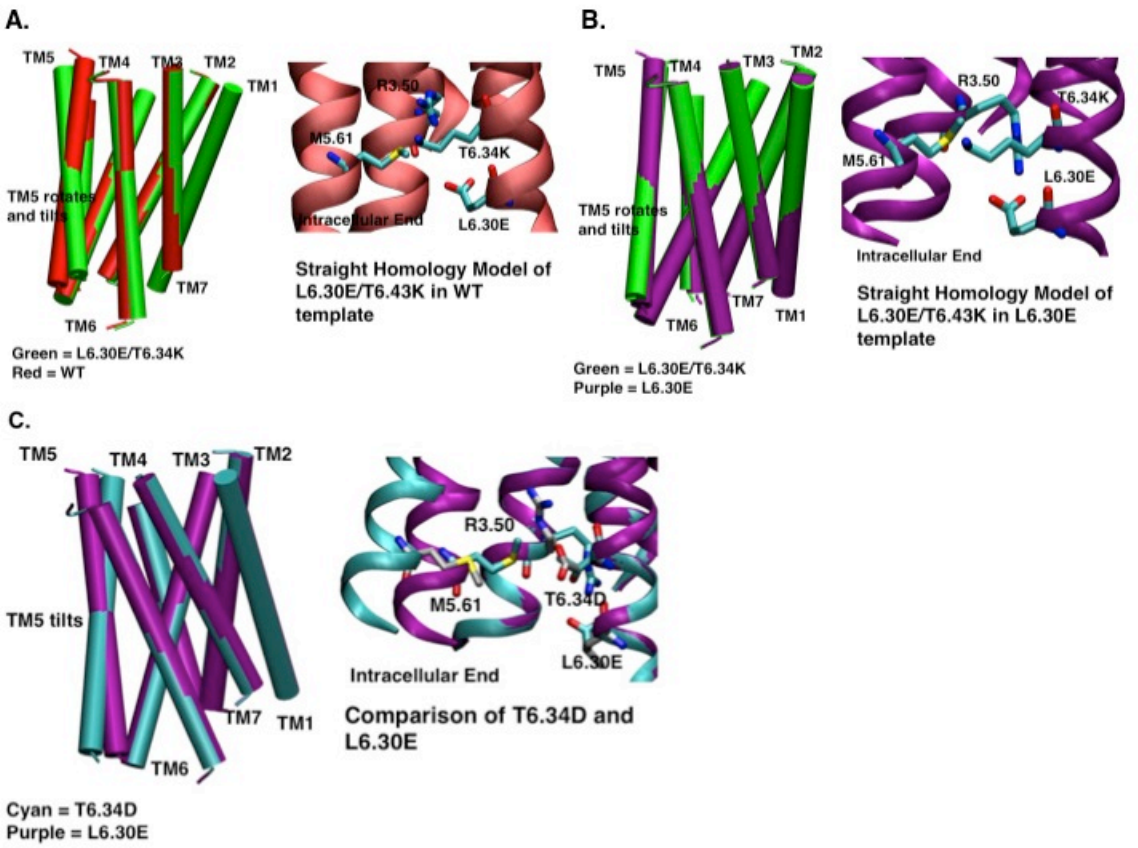


Figure 7.5 Comparison of the predicted structures for the WT (red) and L6.40E/T6.34K (green) mutant **(A)**, L6.40E/T6.34K (green) and L6.30E (purple) mutants **(B)**, and L6.40E (purple) and T6.34D (cyan) mutants **(C)**. Straight homology models of L6.30E/T6.34K in the WT template [WT structure with L6.30E/T6.34K sequence **(A)**] and L6.30E/T6.34K in the L6.30E template [L6.30E/T6.34K structure with L6.30E sequence **(B)**] are shown. An overlap comparison of mutants T6.34D and L6.30E **(C)** are also shown.

In addition, straight homology models of the L6.30E mutant and the L6.30E/T6.34K double mutant are energetically unfavorable (**Figure 7.5B**). By replacing the threonine residue with the lysine one, TM5 tilts and rotates, and the backbone RMSD becomes 1.99 Å. Similarly to the WT and L6.30E/T6.34K double mutant, the straight homology model of the L6.30E/T6.34K sequence in the L6.30E template will not produce energetically favorable structures. By replacing threonine with the long lysine side chain, the R3.50+L6.30E ionic lock cannot form due to the newly introduced interference. T6.34K would clash with M5.61, so TM5 moves to accommodate the new side chain. The WT and L6.30E predicted structures are more similar with a backbone RMSD of 0.80 Å.

The T6.34D and L6.30E receptors are also functionally similar, but structurally different. They have similar binding affinities in the agonist and the GTPγS binding assays, and their backbone

RMSD is rather small at 1.07 Å, but there is significant tilting in TM5 (**Figure 7.5C**). By overlapping the two receptors' structures, we see that if the threonine side chain is increased to the size and shape of the aspartic acid, then it would clash with M5.61. Thus, the single point mutation causes TM5 to swing out to make room for the new side chain. Even though these six receptor structures show significant differences in the positions of TM5, it has no obvious effect on the intracellular hydrogen bond networks involving the DRY motif, the portion of the receptor where the G protein binds. However, TM5 needs to move to provide space for the mutations on TM6. These connections between TM3 and TM6 play important structural, and thereby functional, roles.

CONCLUSIONS

We used the homologized helices from the previously crystallized hCXCR4 GPCR and GEnSeMBLE methodology to predict the structure of hMOR, which was later verified by the crystallized structure of the closely related mMOR. The backbone RMSD of the two receptors is 1.94 Å, well within the crystal resolution of 2.80 Å. The mMOR, crystal structure, however, contains an antagonist β-FNA covalently bound to K5.39, which might affect and constrain the shape of the receptor. We performed sampling on the mMOR crystal structure without the ligand and found five conformations that we predict to be more favorable in energy with backbone RMSDs ranging from 0.52 Å to 1.48 Å with respect to the crystallized mMOR. Not surprisingly, there was great variability in the position and orientation of TM5. These results show that while the crystallized version of mMOR is accurate, it may not be the conformation of the apo-form of the receptor. Finally, using the homologized mMOR helices, we predicted the structures of the hMOR WT receptor and five mutants that have different agonist affinities in the rMOR sequence. The predicted structures contain intracellular interhelical interactions involving the DRY motif that can explain the different levels of activation for the respective receptors. Salt bridges between R3.50 and L6.30E or T6.34D are important for preventing activation. The highly active structure lacks any connections spanning the helical bundle and has a T2.39-R3.50 hydrogen bond, which might stabilize the active conformation. Straight homology models would not be sufficient for predicting structures. Hence, GPCRs like hMOR are flexible and dynamic macromolecules that are sensitive to the structural impact of a single residue. The GEnSeMBLE method is particularly useful for predicting GPCR structures because it performs a thorough sampling of the helical rotation, tilt, and sweep angles and can detect the energetic impact that a single or double point mutation can have on the overall conformation. Since we have accurate

predicted hMOR structures, in the future we will be able to design new ligands that will modulate pain and perhaps be the starting scaffold for new medication.

REFRENCES

1. Waldhoer M, Bartlett SE, Whistler JL (2004) Opioid receptors. Annu Rev Biochem **73**:953-990.
2. Matthes HWD, Maldonado R, Simonin F, Valverde O, Slowe S, Kitchen I, Befort K, Dierich A, LeMeur M, Dolle P, Tzavara E, Hanoune J, Roques BP, Kieffer BL (1996) Loss of morphine-induced analgesia, reward effect and withdrawal symptoms in mice lacking the mu-opioid-receptor gene. Nature **383**:819-823.
3. Raffa RB, Martinez RP, Connelly CD (1994) G-protein antisense oligo deoxyribonucleotides and mu-opioid supraspinal antinociception. Eur J Pharmacol **258**:R5-R7.
4. Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK, Granier S (2012) Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. Nature **485**:321-326.
5. Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, Kobilka BK (2012) Structure of the δ-opioid receptor bound to naltrindole. Nature **485**:400-404.
6. Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, Liu W, Thompson AA, Huang X-P, Carroll FI, Mascarella SW, Westkaemper RB, Mosier PD, Roth BL, Cherezov V, Stevens RC (2012) Structure of the human κ-opioid receptor in complex with JDTic. Nature **485**.
7. Thompson AA, Liu W, Chun E, Katritch V, Wu H, Vardy E, Huang X-P, Trapella C, Guerrini R, Calo G, Roth BL, Cherezov V, Stevens RC (2012) Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. Nature **485**:395-399.
8. Childers SR (1991) Opioid receptor-coupled second messenger systems. Life Sci **48**:1991-2003.
9. Molinari P, Vezzi V, Sbraccia M, Gro C, Riitano D, Ambrosio C, Casella I, Costa T (2010) Morphine-like opiates selectively antagonize receptor-arrestin interactions. J Biol Chem **285**:12522-12535.
10. Ahn KH, Scott CE, Abrol R, Goddard WA, III, Kendall DA (2013) Computationally-predicted CB1 cannabinoid receptor mutants show distinct patterns of salt-bridges that correlate with their level of constitutive activity reflected in G protein coupling levels, thermal stability, and ligand binding. Proteins: Struct Funct Bioinform **81**:1304-1317.
11. Bray JK, Abrol R, Goddard WA, III, Trzaskowski B, Scott CE (2014) SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations. Proc Natl Acad Sci USA **111**:E72-E78.
12. Scott CE, Abrol R, Ahn KH, Kendall DA, Goddard WA, III (2013) Molecular basis for dramatic changes in cannabinoid CB1 G protein-coupled receptor activation upon single and double point mutations. Protein Sci **22**:101-113.
13. Huang P, Visiers I, Weinstein H, Liu-Chen LY (2002) The local environment at the cytoplasmic end of TM6 of the mu opioid receptor differs from those of rhodopsin and monoamine receptors: Introduction of an ionic lock between the cytoplasmic ends of helices 3 and 6 by a L6.30(275)E mutation inactivates the mu opioid receptor and reduces the constitutive activity of its T6.34(279)K mutant. Biochemistry **41**:11972-11980.
14. Abrol R, Bray JK, Goddard WA, III (2012) BiHelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. Proteins: Struct Funct Bioinform **80**:505-518.

15. Abrol R, Griffith AR, Bray JK, Goddard WA, III. Structure prediction of G protein-coupled receptors and their ensemble of functionally important conformations. In: Vaidehi N, Klein-Seetharaman J, Eds. (2012) Membrane Protein Structure: Methods and Protocols. Humana Press, New York, NY.
16. Abrol R, Kim S-K, Bray JK, Griffith AR, Goddard WA, III (2011) Characterizing and predicting the functional and conformational diversity of seven-transmembrane proteins. Methods **55**:405-414.
17. Abrol R, Kim S-K, Bray JK, Trzaskowski B, Goddard WA, III (2013) Conformational ensemble view of G protein-coupled receptors and the effect of mutations and ligand binding. Methods Enzymol **520**:31-48.
18. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res **25**:3389-3402.
19. Katoh K, Kuma K-I, Toh H, Miyata T (2005) MAFFT version 5: Improvement in accuracy of multiple sequence alignment. Nucleic Acids Res **33**:511-518.
20. Wimley WC, Creamer TP, White SH (1996) Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry **35**:5109-5124.
21. http://distill.ucd.ie/porter/.
22. Pollastri G, McLysaght A (2004) Porter: A new, accurate server for protein secondary structure prediction. Bioinformatics **21**:1719-1720.
23. http://www.imtech.res.in/raghava/apssp2/.
24. Raghava GPS. APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. (2002) CASP5. pp. A-132.
25. http://bioinf.cs.ucl.ac.uk/psipred/: 2002-2008.
26. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol **292**:195-202.
27. Kam VWT, Goddard WA, III (2008) Flat-bottom strategy for improved accuracy in protein side chain placements. J Chem Theory Comput **4**:2160-2169.
28. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science **330**:1066-1071.
29. Okada T, Sugihara M, Bondar A-N, Elstner M, Entel P, Buss V (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. J Mol Biol **342**:571-583.
30. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H-J, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. Science **318**:1258-1265.
31. Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman AP, Stevens RC (2008) The 2.6 Å crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. Science **322**:1211-1217.
32. Abrol R, Trzaskowski B, Goddard WA, III, Nesterov A, Olave I, Irons C (2014) Ligand and mutation induced conformational selection in the CCR5 Chemokine G protein-coupled receptor. Under review.
33. Grunbeck A, Huber T, Abrol R, Trzaskowski B, Goddard WA, III, Sakmar TP (2012) Genetically encoded photo-cross-linkers map the binding site of an allosteric drug on a G protein-coupled receptor. ACS Chem Biol **7**:967-972.
34. Berro R, Yasmeen A, Abrol R, Trzaskowski B, Abi-Habib S, Grunbeck A, Lascano D, Goddard WA, III, Klasse PJ, Sakmar TP, Moore JP (2013) Use of G-protein-coupled and

-uncoupled CCR5 receptors by CCR5 inhibitor-resistant and -sensitive human immunodeficiency virus type 1 variants. J Virol **87**:6569-6581.
35. Rasmussen SGF, Choi H-J, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VRP, Sanishvili R, Fischetti RF, Schertler GFX, Weis WI, Kobilka BK (2007) Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. Nature **450**:383-387.
36. Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Li J, Zhang W, Xie X, Yang H, Jiang H, Cherezov V, Liu H, Stevens RC, Zhao Q, Wu B (2013) Structure of the CCR5 chemokine receptor–HIV entry inhibitor Maraviroc complex. Science **341**:1387-1390.
37. Ballesteros JA, Weinstein H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In: Sealfon SC, Ed. (1995) Receptor Molecular Biology. Academic Press, Waltham, Massachusetts, pp. 366-428.
38. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG, Schertler GFX (2008) Structure of a β_1-adrenergic G-protein-coupled receptor. Nature **454**:486-491.
39. Chien EYT, Liu W, Zhao QA, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, Stevens RC (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. Science **330**:1091-1095.
40. Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T, Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. Nature **482**:547-551.
41. Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO, Shaw DE, Weis WI, Wess J, Kobilka BK (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. Nature **482**:552-556.
42. Okada T, Le Trong I, Fox BA, Behnke CA, Stenkamp RE, Palczewski K (2000) X-ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. J Struct Biol **130**:73-80.
43. Teller DC, Okada T, Behnke CA, Palczewski K, Stenkamp RE (2001) Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). Biochemistry **40**:7761-7772.
44. Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y (2002) Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. Proc Natl Acad Sci USA **99**:5982-5987.
45. Li J, Edwards PC, Burghammer M, Villa C, Schertler GFX (2004) Structure of bovine rhodopsin in a trigonal crystal form. J Mol Biol **343**:1409-1438.
46. Standfuss J, Xie G, Edwards PC, Burghammer M, Oprian DD, Schertler GFX (2007) Crystal structure of a thermally stable rhodopsin mutant. J Mol Biol **372**:1179-1188.
47. Stenkamp RE (2008) Alternative models for two crystal structures of bovine rhodopsin. Acta Crystallogr Sect D-Biol Crystallogr **64**:902-904.
48. Murakami M, Kouyama T (2008) Crystal structure of squid rhodopsin. Nature **453**:363-367.
49. Shimamura T, Hiraki K, Takahashi N, Hori T, Ago H, Masuda K, Takio K, Ishiguro M, Miyano M (2008) Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. J Biol Chem **283**:17753-17756.
50. Doré AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, Hurrell E, Bennett K, Congreve M, Magnani F, Tate CG, Weir M, Marshall FH (2011) Structure of the adenosine A_{2A} receptor in complex with ZM241385 and the xanthines XAC and caffeine. Structure **19**:1283-1293.

51. Moukhametzianov R, Warne T, Edwards PC, Serrano-Vega MJ, Leslie AGW, Tate CG, Schertler GFX (2011) Two distinct conformations of helix 6 observed in antagonist-bound structures of a β_1-adrenergic receptor. Proc Natl Acad Sci USA **108**:8228-8232.
52. Yao XJ, Parnot C, Deupi X, Ratnala VRP, Swaminath G, Farrens D, Kobilka B (2006) Coupling ligand structure to specific conformational switches in the β_2-adrenoceptor. Nat Chem Biol **2**:417-422.
53. Ballesteros JA, Jensen AD, Liapakis G, Rasmussen SGF, Shi L, Gether U, Javitch JA (2001) Activation of the β_2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. J Biol Chem **276**:29171-29177.
54. Kobilka BK, Deupi X (2007) Conformational complexity of G-protein-coupled receptors. Trends Pharmacol Sci **28**:397-406.
55. Deupi X, Kobilka BK (2010) Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. Physiology **25**:293-303.
56. Kenakin T, Miller LJ (2010) Seven transmembrane receptors as shapeshifting proteins: The impact of allosteric modulation and functional selectivity on new drug discovery. Pharmacol Rev **62**:265-304.

Chapter VIII

INHIBITION OF BACTERIAL INVASION: AN ANALYSIS OF SMALL LIGAND DOCKING TO THE HUMAN GLYCOPROTEIN ECGP96

Adapted from NIH updates 2011, 2012, and 2013

Done in collaboration with:
* Dr. Tod A. Pascal of the Materials Process and Simulation Center at the California Institute of Technology.
* Professor Prasadarao V. Nemani, Dr. Subramanian Krishnan, and Dr. Muthusamy V. Shanmuganathan of the Division of Infectious Diseases, Department of Pediatrics, and Department of Surgery, Children's Hospital Los Angeles, and University of Southern California, Los Angeles
* Professor Brian M. Stoltz and Dr. Douglas Behenna of the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena

ABSTRACT

The bacterium *Escherichia coli* penetrates the blood brain barrier via sugar moiety chitobiose-mediated interactions with the human glycoprotein Ecgp96 leading to the onset of the potentially lethal neonatal meningitis. By blocking Ecgp96's glycosylation site with small molecules, chitobiose binding to bacteria and subsequent bacterial invasion can be prevented. Experimental studies show that three ligands, Doxazosin, MSR15, and Telmisartan, are effective at inhibiting bacterial invasion. We use hierarchical docking methods, DarwinDock and GenDock, to dock these ligands to Ecgp96 and based on the predicted binding sites, we suggest modifications of the three ligands that we predict will be stronger binders and would be more effective at preventing bacterial invasion. Furthermore, we dock the small molecules Geldanamycin and Z154, one of the identified strong binders, to the glycosylated form of Ecgp96 and identify two other potential regions that ligands would bind to. We ran virtual ligand screening on the Geldanamycin binding site to find novel ligands that could bind to unexplored regions of Ecgp96. Three of the 500 ligands selected by DOCK Blaster 3 form three hydrogen bonds with the protein. The new ligands from this study are shared with our experimental collaborators, so that they can synthesize them (if needed) and determine their binding affinities and effectiveness at preventing meningitis.

INTRODUCTION

Neonatal meningitis is a serious disease affecting the central nervous system, resulting in hearing loss, blindness, convulsions, abnormal speech, mental retardation, and often death. Lack of knowledge about the mechanism of disease onset has inhibited the development of vaccines and treatment [for reviews, see (1-4)]. Previous studies from the Nemani and Goddard labs have elucidated the process of bacterial invasion causing this disease (5-8). *Escherichia coli* K1 is transported across the blood-brain barrier (BBB) via sugar moiety-mediated interactions between *E. coli's* outer membrane protein A (OmpA) and the human glycoprotein Ecgp96 (5). Ecgp96, a receptor with a large sequence identity with the previously crystallized heat shock protein 90 (Hsp90), is located on the surface of the human brain microvascular cells (HBMCs) (6). The sugar moiety N-acetyl-D-glucosaminyl-β(1-4)-N-acetyl-D-glucosamine (chitobiose) binds to one of three glycosylation sites on the protein (including N142 and N252 in the model discussed below), facilitating interactions between Ecgp96 and OmpA. Recently published work from the Goddard and Nemani labs indicate that the chitobiose epitopes bind to the loops of OmpA (7). Ecgp96 has been experimentally proven to associate with the angiotensin II type 1 receptor

(AT1R), and it is hypothesized that this molecular complex governs protein kinase C-α (PKC-α) phosphorylation during *E. coli* invasion (9).

Currently, we are interested in finding where chitobiose interacts with Ecgp96. Previously, Dr. Tod Pascal and co-workers of the Goddard lab computationally docked 52 ligands that had been crystallized with Hsp90 and used the 15 lowest energy binders with the software program Phase to produce a pharmacophore of the Hsp90 binding site. Virtual ligand screening (VLS) identified 2000 small molecules out of a database of 2 million ligands that at least partially matched this proposed pharmacophore. Fifty ligands, 45 of which exactly matched the pharmacophore, were selected to dock to Ecgp96. Prof. Nemani's lab tested the Goddard lab's ligand recommendations as well as others synthesized by a collaborator with the pharmacophore in mind.

Table 8.1 shows three of the seven ligands that were effective in preventing bacterial invasion: MSR15, Doxazosin, and Telmisartan with the latter two in the list of 50 ligands obtained above. The inhibition percentage in **Table 8.1** ranges from 90.8% with Doxazosin to 74.5% with Telmisartan. Doxazosin, an alpha-blocker that treats high blood pressure, and Telmisartan, an AT1R blocker that treats hypertension, are both commercially available. Telmisartan is especially interesting because of when it binds to the AT1R, which interacts with Ecgp96, the drug prevents PKC-α phosphorylation and subsequent bacterial invasion and onset of meningitis in mouse models (9). Also, the AT1R blocker Candesartan has been shown to bind directly to the Ecgp96 homologue Hsp90 (10), which suggests Telmisartan might bind to Ecgp96 as well as the AT1R in the protein complex (9). MSR15 was specially synthesized for this study. Here, we dock these three ligands discussed above to identify their respective binding sites, and then we offer suggestions for new derivatives with enhanced affinity for interacting with Ecgp96. We expect that these ligands will be more effective in inhibiting bacterial invasion.

Table 8.1 Percentage of *E. coli* inhibition caused by ligands matching a pharmacophore generated based on computational docking studies of known ligands to similar protein Hsp90.

Ligand	Average Bacterial Inhibition
Doxazosin	90.8%
MSR15	78.7%
Telmisartan	74.5%

As noted above, Hsp90 and Ecgp96 contain glycosylation sites. However, Hsp90 has not been crystallized with a ligand in its glycosylated form [for example, see (11)], so it is unknown where co-crystallized molecules like Geldanamycin would bind in the presence of the glycans. We

consider this possibility by docking Geldanamycin to five different regions of the glycosylated Ecgp96 protein and find that the ligand has more favorable binding energies with other regions of the protein if the glycosylated unit is present in the traditional ATP binding site. Since we have identified possible novel binding sites that have not been explored before, we conduct virtual ligand screening (VLS) to identify small molecules that could preferentially bind to this portion of the protein that is accessible in the biologically relevant glycosylated form.

METHODS

(For more details about docking methodology, see *Chapter II*.)

Preparing the protein Ecgp96 model

Previous work from the Goddard lab suggests that Ecgp96 has one transmembrane region and a solvent-exposed globular portion, which contains two of the three glycosylation sites where chitobiose can interact (N142 and N252). One of the sites, residue numbers 142-145 with a sequence of NASD, is located in the ATP binding site. In previous work, a model of human Ecgp96 protein residue numbers 102 to 618 was constructed using a homology model based on the Hsp90 crystal structure due to the two proteins' high sequence identity. Pascal *et al.* used Swiss Model software to generate the structure of the globular portion of Ecgp96 consisting of 191 residues based on the Hsp90 crystal structure. Spheres were generated for this SGB (12) solvated version of this predicted protein structure. The actual protein used for docking has CHARMM (13) charges and neutralized terminal ends and was not minimized in a solvated environment, but since its heteroatom RMSD is 0.195 Å with respect to the solvated protein, the sphere sets for the two proteins should not differ significantly. Doxazosin, Telmisartan, and MSR15 were docked to the model of protein before minimization with SGB solvation. Geldanamycin was docked to the same model after SGB solvated minimization. Spheres were regenerated for the protein after minimization.

To create the glycosylated version of the Ecgp96 protein, we abstracted glycosylated sugars from the crystal structure of human glucosylceramidase (PDB ID: 2V3F) (14) and linked them to either N142 and N252, both Ecgp96 glycosylation sites, of the Ecgp model after minimization with SGB solvation. The N-linked glycan consists of two N-acetyl-D-glucosamine monosaccharides and a terminal alpha-D-mannose sugar.

In order to locate binding sites for various ligands, we generated five sphere regions each for the non-minimized and minimized forms of the protein. Two sphere sets for the minimized protein are indicated in **Figure 8.1**. Region 1, in yellow, is particularly important because it contains one of the glycosylation sites at residue N142. This is the conserved residue that Geldanamycin (11) and ATP (15) bind to in the crystallized Hsp90 proteins. Ligands docked to Region 2, in purple, form favorable or low energy complexes with the glycosylated protein that will be discussed below.

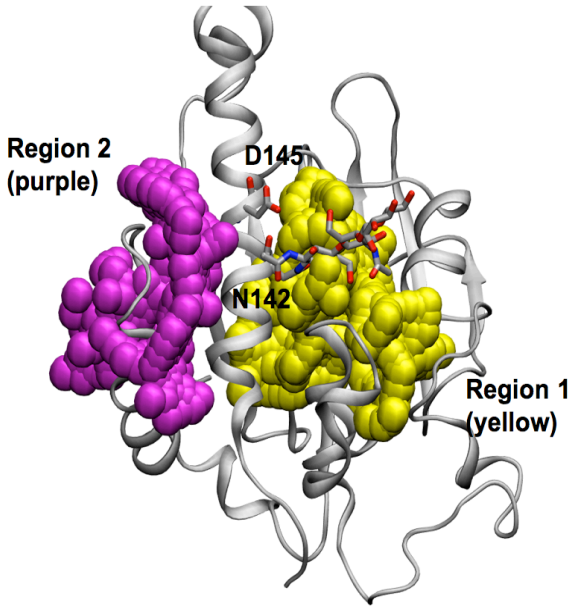


Figure 8.1 Two of the five regions of the solvated minimized human Ecgp96 protein to which small molecule ligands were docked to. Region 1 (yellow) contains the glycosylation site involving residues N142 and D145. Region 2 (purple) does not cover any glycosylation site, but ligands have strong binding energies in this space on the glycosylated protein.

Constructing small molecule ligands

Ligands were constructed with Maestro software (16) and used Jaguar software (17) to determine Mulliken charges. A MacroModel (18; 19) conformational search was performed as described in *Chapter II*. Two hundred and forty-four poses of Doxazosin, 825 poses of MSR15, and 3309 poses of Telmisartan were minimized with SGB solvation (12) for 100 steps. Four hundred and sixty-two poses of Z154 were minimized with SGB solvation for 200 steps or to the RMS threshold of 0.1 kcal/mol/Å. The poses of Doxazosin and Telmisartan were clustered twice with a diversity of 2.0 Å, whereas the poses of MSR15 were clustered by 2.0 Å, then 2.5 Å, and the poses of Z154 were clustered for two rounds, each with a diversity of 1.0 Å. Ten conformations of Doxazosin, 10 conformations of MSR15, 11 conformations of Telmisartan, and 17

conformations of Z154 were used for docking. After clustering, the ligand poses are minimized 100 steps with MPSim software (20) and SGB solvation.

Geldanamycin, one of the ligands used for docking, was extracted from the crystallized Hsp90 (PDB ID: 1YET) (11). Geldanamycin was minimized 361 steps to an RMS threshold of 0.2 kcal/mol/Å using MPSim (20) software with SGB solvation and the Dreiding III (21) force field with the Lennard-Jones van der Waals (vdw) potential.

Docking small molecules to Ecgp96

To determine how these three ligands interact with the Ecgp96, we used DarwinDock and GenDock hierarchical docking programs (22-26). The DarwinDock program generated 5000 poses with DOCK 6 (27) of the ligand in the alanized protein, clustered similar poses within a specified diversity of 2.0 Å into families, and generated 5000 new conformations. This cycle was repeated until no more new conformations could be generated, or no more than 2% of the current number of families could be created. Overall approximately 50,000 poses were generated and clustered into ~1500 families. A family head, average representative of each family, was chosen and the respective energies were calculated. The 10% of family heads with lowest energies were selected, and the energies were evaluated for all the members of these families. From these poses, fifty poses were chosen with the lowest polar energy, fifty were chosen with the lowest phobic energy, and fifty were chosen with the lowest total energy.

The next step was to refine the selected docked poses. The bulky hydrophobic groups replaced the smaller alanine residues, and simultaneously polar residues in the binding sites were optimized using the program Side chain Rotamer Excitation Method analysis (SCREAM) (28). The complex was minimized 10 steps. Then the charged residues (D, E, H, K, R) were neutralized because electrostatic interactions can have a large impact on energy calculations, and, again, the complex was minimized by 10 steps. Finally, the entire complex was minimized 100 steps and ranked according to various energy categories including the local cavity energy, binding energy, full Delphi energy, partial Delphi energy, total energy, and unified cavity energy. Local cavity energy is the interaction between the ligand and the residues in the binding sites. The binding energy is the equivalent of the protein and ligand energies subtracted from the complex energy. The full Delphi energy is the energy of the solvated complex, and the partial Delphi energy is the energy with only the ligand solvated. The unified cavity energy equals the energy of the ligand and the residues of the binding site of any ligand pose generated.

Neutral versions of Doxazosin, MSR15, and Telmisartan were docked to the non-glycosylated form of Ecgp96. Geldanamycin was docked to both the non-glycosylated and the glycosylated versions of Ecgp96. Z154 was docked to only the glycosylated form of this protein.

Matching novel ligands into the binding site of docked ligands

The ultimate goal of this study was to identify new ligands that have strong binding affinities for Ecgp96 and prevent chitobiose binding. The online database PubChem (29) was used to search for known small molecules. The online server ALOGPS 2.1 (30) calculated the solubility values of the ligands. To verify the effectiveness of binding, we matched the derivatives to the original ligand pose of the docked complex. The generated charged ligand pose was extracted from the complex generated during the initial step DarwinDock. In Maestro, it was altered to resemble the derivative. The resulting ligand was minimized in MacroModel with the solvent specified and dielectric constant specified for water. The OPLS-2005 force field (31) was used for minimization. Mulliken charges were calculated with Jaguar software. The new ligand was inserted into the respective alanized version of Ecgp96 matching the protein the original ligands were docked to. Subsequently, all the original residues replaced the alanines, and residues in the binding site were optimized using the SCREAM program. These residues were specified in the second step of the original docking procedure of DarwinDock. Then, selected individual residues in the binding site were optimized with SCREAM in order to maximize their interactions with the ligand. Forty rotamers were produced. All 41 complexes were minimized by 10 steps in vacuum with the MPSim program (20) and the universal force field Dreiding III (21). These complexes were neutralized and completely minimized using the last two modules in GenDock.

VLS with DOCK Blaster 3 to identify novel ligands to bind to the glycosylated Ecgp96

After docking Geldanamycin to our model of the glycosylated human glycoprotein Ecgp96, we identified the lowest energy binding site for the ligand (**Figure 8.1)**, which is different from the non-glycosylated Geldanamycin binding site. We performed VLS to identify possible ligands that would bind to this new site when the protein is glycosylated.

We took the glycosylated Ecgp96-Geldanamycin complex with the lowest binding and cavity energies with strain and ligand solvation, removed the ligand, and selected the residues that were within 6 Å of the docked Geldanamycin. The glycosylated Ecgp96 protein was our target, and the selected residues were the binding site provided to DOCK Blaster 3 (32). DOCK Blaster

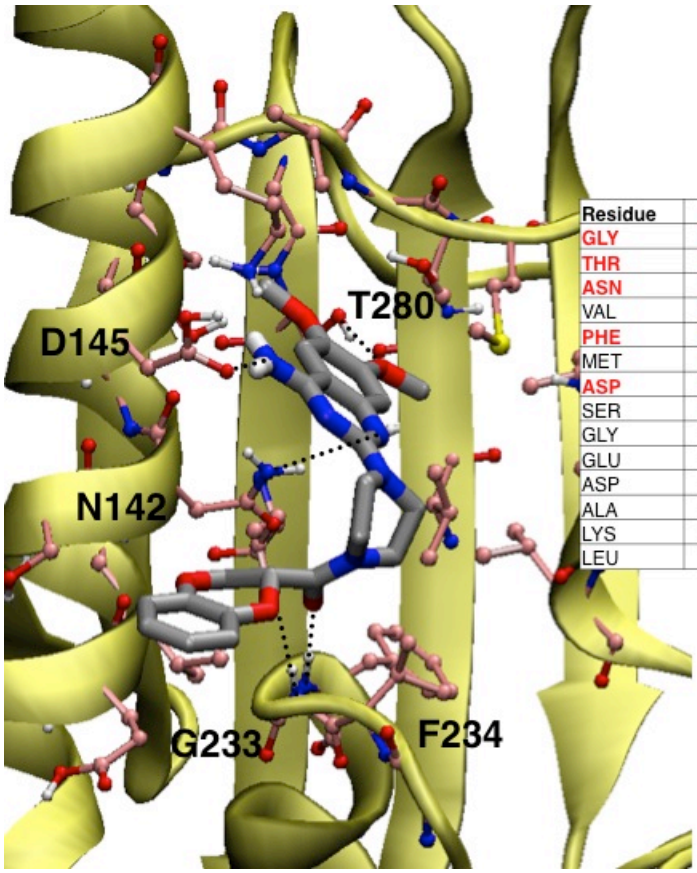
docked small molecules from a database to the provided structure and ranked them according to energy. The structures of 500 lowest energy ligands, or hits, were provided. We chose 19 low energy hits with significant structural differences.

We enriched and evaluated the complexes using a force field and program, GenDock, created in our lab. First, we calculated the Mulliken charges for the ligands using Maestro (16) and Jaguar (17) software by performing a single energy point calculation with density functional theory (DFT), the B3LYP functional, and the 6-31G** basis set. The ligands were merged with an alanized form of the glycosylated Ecgp96 protein. By 'alanized,' we mean that the large, bulky hydrophobic residues (M, F, I, L, Y, V, W) were replaced with alanine residues. Then, we used the SCREAM program (28) to replace the alanines with the correct sequence and to sample different rotamers of all the residues within 4 Å of the ligand before minimizing the entire complex for 10 steps. The next step was to neutralize the protein by removing protons from the basic residues (R, K, and H) and adding protons to the acidic residues (D and E). Lastly, we minimized the complex by 110 steps and determined the binding energy. We used the Dreiding III force field (21) with the exponential-6 term replaced with the Lennard-Jones term (DIIIFF-LJ) to evaluate energies.

RESULTS AND DISCUSSION

Doxazosin derivatives can attach fluorophores to track Ecgp96 at the BBB

Previous studies indicate that Doxazosin is the most effective in preventing bacterial invasion, as seen in **Table 8.1**. **Figure 8.2** shows our docking results, where complexes were selected based on their lowest energy ranking in the most number of categories including binding energy, binding energy with ligand strain, binding energy with Delphi solvation and ligand strain, full Delphi, partial Delphi, and total energy. Doxazosin forms five hydrogen bonds, two with residues N142 and D145, which are located in one of the glycosylation sites. Thus, our results agree with experimental data because Doxazosin can prevent chitobiose from interacting with the glycosylation site. Another two hydrogen bonds are formed with backbone amide protons of G233 and F234, so mutations would not affect the binding of Doxazosin or its effectiveness in preventing bacterial invasion. The fifth hydrogen bond is formed with T280. The ligand also has hydrophobic interactions with V232 and M189.



Residue		vdW	Coulomb	H-Bond	NonBond
GLY	**233**	**0.273**	**-2.332**	**-3.687**	**-5.747**
THR	**280**	**2.715**	**-3.24**	**-4.577**	**-5.102**
ASN	**142**	**-3.267**	**-0.271**	**-1.319**	**-4.858**
VAL	232	-3.821	-1.012	0	-4.832
PHE	**234**	**3.085**	**-2.791**	**-4.724**	**-4.43**
MET	189	-3.947	0.034	0	-3.912
ASP	**145**	**2.75**	**-0.729**	**-5.08**	**-3.059**
SER	141	-1.411	-0.965	0	-2.377
GLY	188	-1.198	-0.8	0	-1.999
GLU	138	-1.986	0.067	0	-1.919
ASP	184	-0.639	-0.578	0	-1.217
ALA	143	-0.761	-0.415	0	-1.176
LYS	149	-0.483	-0.656	-0.006	-1.144
LEU	198	-1.135	0.061	0	-1.074

Figure 8.2 Doxazosin docked to human Ecgp96 using GenDock. This Doxazosin-Ecgp96 pose has the lowest binding energy, binding energy with strain, binding energy with delphi solvation and strain, full delphi, partial delphi, and total energy. The prediction shows that Doxazosin forms five hydrogen bonds with T280, D145, N142, and with the backbone amide protons of G233 and F234. D145 and N142 are residues in one of the glycosylation sites, so our results indicate that Doxazosin would be effective in blocking that region from interactions with chitobiose, which agrees with experimental data. Hydrogen bonds are indicated by black dotted lines in the pictures and are in bold red in the tables. The cavity analysis breaks down the non-bonding energy into van der Waals (vdW), coulombic (Coulomb), and hydrogen bond (H-Bond) interactions. Only residues whose energetic interactions are greater than 1 kcal/mol are shown in the table.

In future studies, our experimental collaborators will attach a fluorophore to a ligand to monitor Ecgp96's behavior after binding to the small molecule. Our proposed Doxazosin derivatives, which are shown in **Figures 8.3B-D**, are functionalized with a ketone, for example Pubchem compound ID (CID) 15591670 (S4.2) and CID15591672 (S4.3), or a sulfonamide, such as CID15591676 (S4.4), so that they can react and attach to the fluorophore. Matching results, or altering the docked ligand conformation within the protein to resemble the derivative, show that these functional groups are exposed to the solvent and in turn suggest that the attached fluorophore would not disrupt the ligands' binding to Ecgp96.

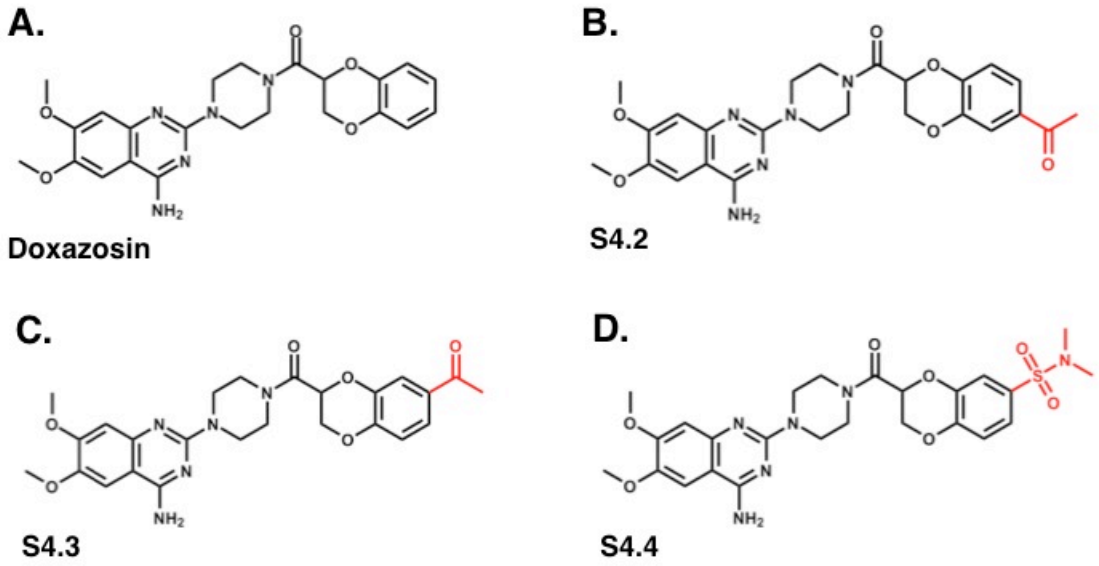


Figure 8.3 Schematic drawings of **(A)** Doxazosin and the proposed derivatives including **(B)** Pubchem compound ID (CID) 15591670 (S4.2), **(C)** CID15591672 (S4.3), and **(D)** CID15591676 (S4.4). Red portions of the ligand are parts altered from the Doxazosin scaffold.

Furthermore, the energy results of matching the derivatives to Doxazosin show that the calculated interactions energies will decrease, or become more favorable, as seen in **Table 8.2**. For example, CID15591670 (S4.2) is significantly lower in local cavity energy, binding energy, full Delphi energy, and unified cavity energy. Lower binding and cavity energies suggest that these derivatives have a stronger interaction with Ecgp96 and could possibly be more effective than Doxazosin in preventing chitobiose-mediated interactions with *E. coli* at the BBB.

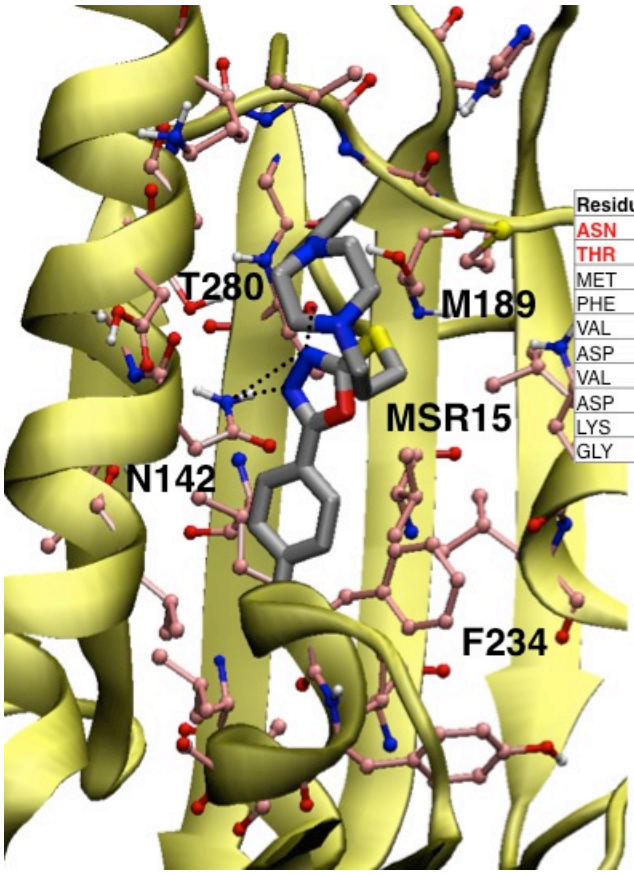
Table 8.2 Energy results of matching the Doxazosin derivative ligands shown in **Figure 8.2B-D** to the Doxazosin-Ecgp-96 complex shown in **Figure 8.1**. All the derivatives are lower than or comparable to Doxazosin in various energy types and are significantly lower in terms of unified cavity energy. In these complex poses, the red portions of the ligands are exposed and can be functionalized by fluorescent tags.

Ligand	Local Cavity Energy (kcal/mol)	Binding Energy (kcal/mol)	Full Delphi (kcal/mol)	Partial Delphi (kcal/mol)	Total Energy (kcal/mol)	Unified Cavity Energy (kcal/mol)
Doxazosin	-45.56	-69.88	-29.17	-47.75	-869.80	-41.17
15591670 (S4.2)	-47.75	-71.78	-34.48	-48.38	-831.86	-47.75
15591672 (S4.3)	-48.85	-73.48	-37.27	-50.25	-839.02	-48.85
15591676 (S4.4)	-49.95	-74.08	-35.35	-48.05	-765.87	-49.95

MSR15 blocks the glycosylation site and chitobiose interactions

MSR15 is the only ligand discussed here that is not commercially available since it was synthesized by Prof. Nemani's collaborators. It is the second-most effective ligand in blocking bacterial invasion because it was successful in doing so for 78.7% of all cases. **Figure 8.4** shows

our proposed binding site generated from docking MSR15 to Ecgp96. MSR15 forms four hydrogen bonds with two residues, N142 and T280. N142 is one of the residues in the glycosylation site, so our results support experimental data because we prove that MSR15 is effective in blocking the site from chitobiose.



Residue		vdW	Coulomb	H-Bond	NonBond
ASN	142	0.092	-2.431	-6.847	-9.186
THR	280	2.691	-3.043	-5.91	-6.263
MET	189	-4.196	0.137	0	-4.059
PHE	234	-1.779	-0.107	0	-1.886
VAL	187	-0.878	-0.633	0	-1.511
ASP	145	-1.451	0.215	0	-1.235
VAL	246	-1.086	-0.13	0	-1.216
ASP	184	-0.907	-0.265	0	-1.172
LYS	149	-0.936	-0.163	0	-1.099
GLY	186	-0.535	-0.491	0	-1.026

Figure 8.4 MSR15 docked to human Ecgp96 using GenDock. This MSR15-Ecgp96 pose has the lowest binding energy, partial delphi, and unified cavity energy. The prediction shows that MSR15 forms four hydrogen bonds with T280 and N142. N142 is in one of the glycosylation sites, so our prediction indicates that MSR15 can block chitobiose from interacting with the protein, which agrees with experimental data. Hydrogen bonds are indicated by black dotted lines in the pictures and are in bold red in the tables. The cavity analysis breaks down the non-bonding energy into van der Waals (vdW), coulombic (Coulomb), and hydrogen bond (H-Bond) interactions. Only residues whose energetic interactions are greater than 1 kcal/mol are shown in the table.

Based on the binding site in **Figure 8.4**, we identified several derivatives of MSR15 as shown in **Figures 8.5B-G**. Ligands in **Figures 8.5B-E** were found in a search of the database PubChem (29). The other two ligands in **Figures 8.5F** and **8.5G** were designed based on the binding site. Ligands in **Figures 8.5B-D** are commercially available and can be purchased from from the vendor Ambinter, while the other ligands need to be synthesized.

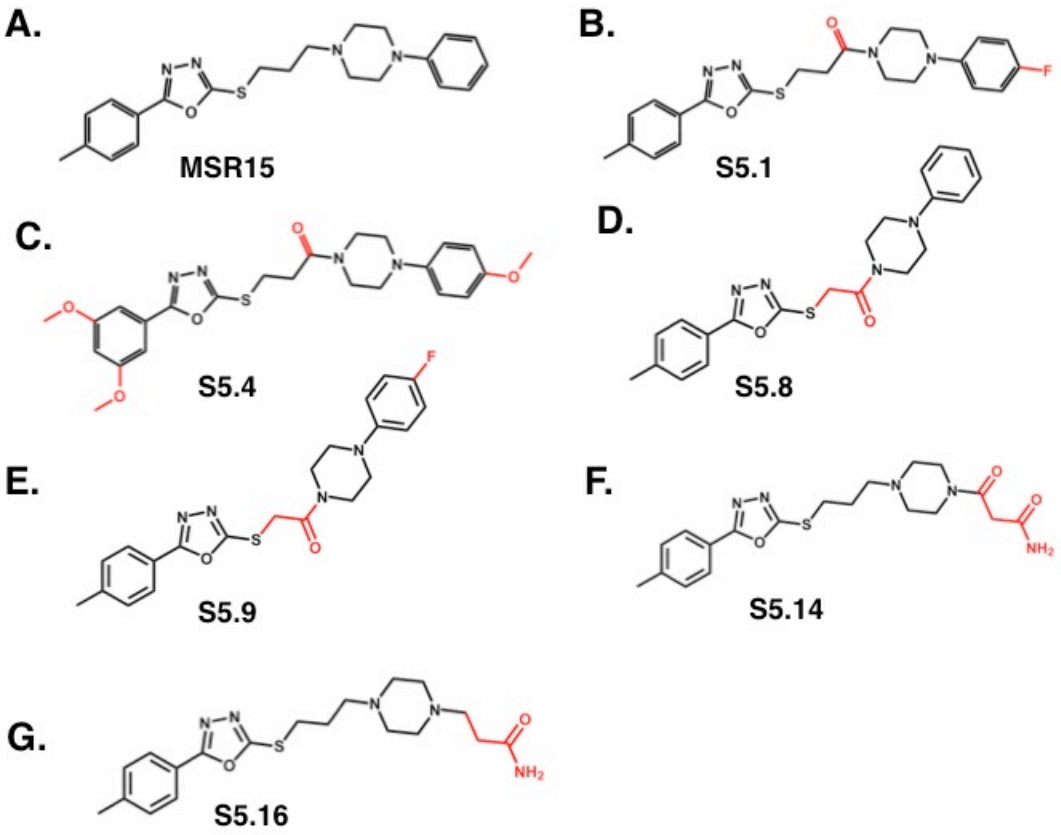


Figure 8.5 Schematic drawings of **(A)** MSR15 and the proposed derivatives including **(B)** Ambinter ID (AID) 81805868 (S5.1), **(C)** AID82150035 (S5.4), **(D)** AID83257400 (S5.8), **(E)** CID7379308 (S5.9), and our own designs **(F)** S5.14 and **(G)** S5.16. Red portions of the ligand are parts altered from the MSR15 scaffold. Ligands **B-D** are commercially available from Ambinter. Ligands **E-G** need to be synthesized.

Our results from matching the ligand derivatives to MSR15 in the Ecgp96 complex are shown in **Table 8.3**. All of the ligands included are lower (more favorable) or comparable in energy to MSR15. Ligand S5.14 shows the most improvement because its complex's unified cavity energy is 10 kcal/mol lower and its full Delphi energy is 30 kcal/mol lower than that of the MSR15 complex. Complexes with derivatives AID82150035 (S5.4) and AID83257400 (S5.8) do not show a significant improvement, but they are included here since the ligands are commercially available and are easier to purchase than MSR15.

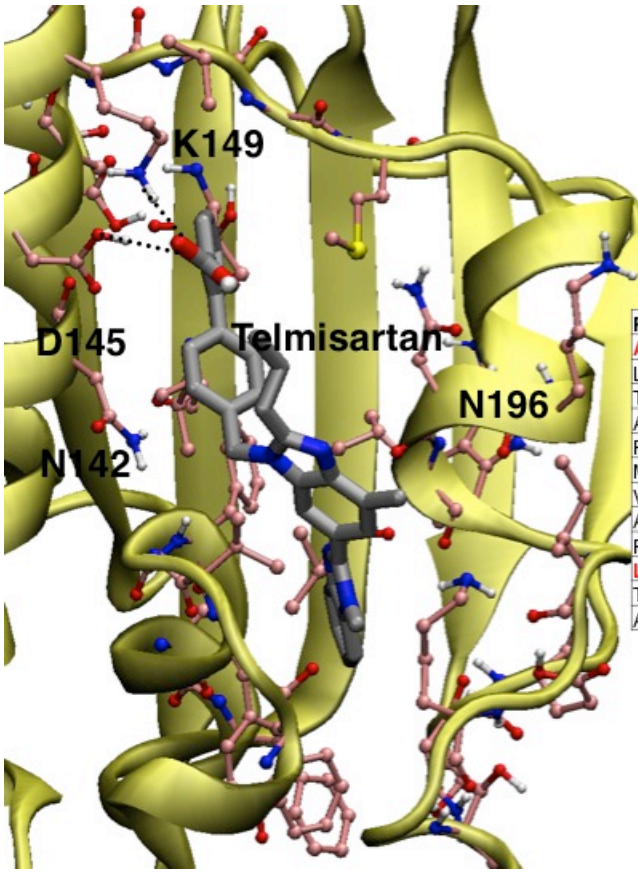
Table 8.3 Energy results of matching the MSR15 derivative ligands shown in **Figure 8.5B-F** to the MSR15-Ecgp96 complex shown in **Figure 8.3**. All the derivatives are lower than or comparable to MSR15 in various energy types.

Ligand	Local Cavity Energy (kcal/mol)	Binding Energy (kcal/mol)	Full Delphi Energy (kcal/mol)	Partial Delphi Energy (kcal/mol)	Total Energy (kcal/mol)	Unified Cavity Energy (kcal/mol)
MSR15	-28.64	-53.75	-33.24	-45.17	-766.97	-30.48
81805868 (S5.1)	-34.53	-56.42	-34.28	-44.88	-760.83	-36.38
82150035 (S5.4)	-30.01	-53.64	-19.98	-35.25	-776.77	-31.03
83257400 (S5.8)	-28.73	-54.72	-31.28	-42.77	-812.04	-31.60
7379308 (S5.9)	-34.76	-64.32	-41.61	-51.13	-822.92	-34.76
S5.14	-38.68	-77.74	-63.53	-48.84	-871.81	-40.64
S5.16	-32.92	-69.76	-46.20	-55.86	-818.96	-34.49

Telmisartan has strong binding energies for the Ecgp96 protein

Unlike Doxazosin and MSR15, we have some knowledge about how Telmisartan binding impacts the bacterial invasion mechanism. While it has been shown that Telmisartan interacts with AT1R that is in complex with Ecgp96 (9), we show here that Telmisartan has strong binding energies with the Ecgp96 protein itself and suggest that it may bind to either protein. Previously, another AT1R blocker has been proven to bind directly to the similar Hsp90 protein (10), so it is likely that Telmisartan can bind to Ecgp96. Telmisartan is effective in preventing bacterial invasion with a 74.5% success rate, but its usefulness is hampered by its low solubility value. The ALOGPS 2.1 predicted solubility of Telmisartan is 3.50 mg/l. This concentration is relatively low compared with Doxazosin whose solubility is predicted to be 790 mg/l. Our goal is to develop a ligand that has an equally strong binding affinity for Ecgp96 as Telmisartan, but is more soluble.

Figure 8.6 shows the lowest local cavity and unified cavity energy product from docking Telmisartan to Ecgp96. Telmisartan forms two hydrogen bonds with D145 and K149. Our docking results support experimental data because D145 is part of one of the glycosylation sites. Telmisartan also has significant non-bonding interactions with N142, another residue in the glycosylation site. Therefore, Telmisartan blocks chitobiose from interacting with Ecgp96. Based on this proposed binding pose, derivatives of Telmisartan were designed to maximize favorable interactions with surrounding residues as well as improve the solubility.



Residue		vdW	Coulomb	H-Bond	NonBond
ASP	**145**	**2.175**	**-3.444**	**-4.52**	**-5.789**
LEU	198	-5.187	0.284	0	-4.903
TYR	235	-3.365	-0.179	0	-3.545
ASN	142	-3.296	0.013	0	-3.283
PHE	234	-3.35	0.07	0	-3.28
MET	189	-2.954	-0.116	0	-3.069
VAL	232	-3.169	0.324	0	-2.845
ASN	197	-3.04	0.45	0	-2.591
PHE	230	-1.972	-0.142	0	-2.114
LYS	**149**	**5.081**	**-1.703**	**-5.25**	**-1.873**
THR	280	-0.772	-0.477	0	-1.25
ALA	146	-1.201	0.123	0	-1.078

Figure 8.6 Telmisartan docked to human Ecgp96 using GenDock. This Telmisartan-Ecgp96 pose has the lowest local cavity energy and unified cavity energy. The prediction shows that telmisartan forms two hydrogen bonds with D145 and K149. D145 is a residue in one of the glycosylation sites, so our results indicate that Telmisartan would be effective in blocking that region from chitobiose interactions, which agrees with experimental data. Hydrogen bonds are indicated by black dotted lines in the pictures and are in bold red in the tables. The cavity analysis breaks down the non-bonding energy into van der Waals (vdW), coulombic (Coulomb), and hydrogen bond (H-Bond) interactions. Only residues whose energetic interactions are greater than 1 kcal/mol are shown in the table.

Figure 8.7B shows TelmisartanB, a ligand synthesized in Prof. Brian Stoltz's lab and tested for effectiveness by Prof. Nemani's group (9). We predict that its solubility value is 4.70 mg/l, which is a slight improvement over Telmisartan (see **Table 8.4**). Furthermore, it is experimentally proven to be equally effective at preventing the onset of meningitis in mouse models (9). **Table 8.5** shows a comparison of our calculated energies for matched poses with the original docked Telmisartan-Ecgp96 complex. TelmisartanB is at least 7 kcal/mol lower in energy in every energy category except total energy. Our docking results support experimental observations because we have shown that TelmisartanB has a strong affinity for interacting with Ecgp96.

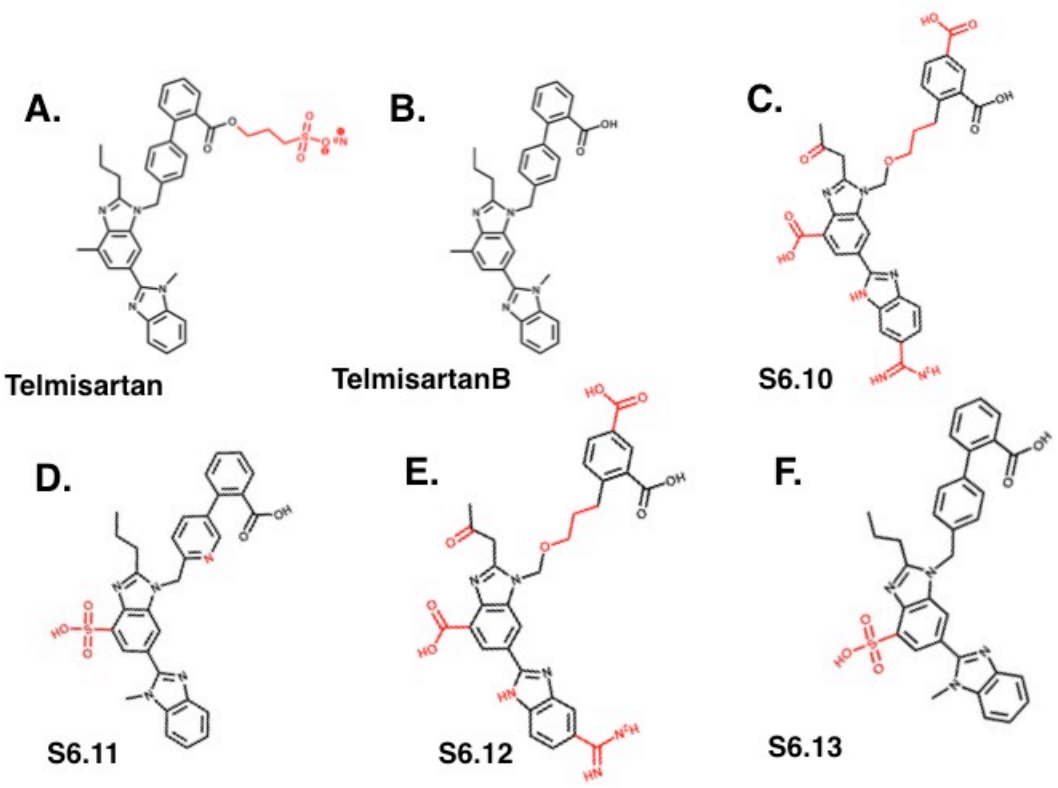


Figure 8.7 Schematic drawings of **(A)** Telmisartan and the proposed derivatives including **(B)** TelmisartanB, **(C)** S6.10, **(D)** S6.11, **(E)** S6.12, and **(F)** S6.13. Red portions of the ligand are parts altered from the Telmisartan scaffold. Ligand **B** was designed and synthesized by Prof. Brian Stoltz and co-workers, and has been experimentally shown to be as effective as Telmisartan in inhibiting bacterial invasion. Ligands **C-G** were designed by the Goddard group members and need to be synthesized by collaborators. All the derivatives have higher solubility than Telmisartan.

Table 8.4 Calculated ALOGPS 2.1 solubility values of Telmisartan shown in **Figure 8.7A** and the Telmisartan derivative ligands shown in **Figure 8.7B-F**. All the derivatives are two to ten times lower than Telmisartan in solubility.

Ligand	Solubility (mg/l)
Telmisartan	3.50
TelmisartanB	4.71
S6.10	30.15
S6.11	14.23
S6.12	29.43
S6.13	7.67

Table 8.5 Energy results of matching the telmisartan derivative ligands shown in **Figure 8.7B-F** to the Telmisartan-Ecgp96 complex shown in **Figure 8.6.** All the derivatives are lower than or comparable to Telmisartan in various energy types. TelmisartanB is significantly lower in local cavity energy, binding energy, full Delphi energy, partial delphi energy, and unified cavity energy. Experimental evidence supports our matching results because experiments indicate that TelmisartanB is equally effective as Telmisartan in preventing bacterial invasion.

Ligand	Local Cavity Energy (kcal/mol)	Binding Energy (kcal/mol)	Full Delphi (kcal/mol)	Partial Delphi (kcal/mol)	Total Energy (kcal/mol)	Unified Cavity Energy (kcal/mol)
Telmisartan	-44.42	-63.09	-27.06	-50.81	-746.42	-45.44
TelmisartanB	-57.07	-78.37	-34.76	-59.40	-734.02	-57.72
S6.10	-56.36	-86.77	-51.65	-57.54	-885.10	-56.60
S6.11	-46.32	-62.69	-22.32	-43.43	-754.93	-46.87
S6.12	-61.67	-87.26	-52.31	-57.75	-870.38	-61.93
S6.13	-45.72	-62.96	-26.11	-44.54	-754.92	-46.24

Figures 8.7C-F are Telmisartan derivatives that we designed based on our predicted binding site. A solubility comparison is shown in **Table 8.4**. Our proposed ligands have improved solubility values that are two to ten times larger than that of Telmisartan. **Table 8.4** shows that not only do the ligands have better solubility, but more affinity for the protein. All the ligands are lower in local cavity, total, and unified cavity energies. Ligand S6.12's pose shows the largest improvement in energies over Telmisartan. For example, its unified cavity energy is 16.5 kcal/mol lower in energy, and its binding energy is 14 kcal/mol lower in energy than the docked Telmisartan pose. The next step is to dock these poses to see if there are any lower energy binding sites for the ligands.

Geldanamycin would preferentially bind to a different site depending upon glycosylation

To test the impact of glycosylation on known ligand binding, we docked Geldanamycin to unglycosylated forms of Hsp90 and Ecgp96 to compare with the crystallized version of the Hsp90-Geldanamycin complex. Upon confirming that we can accurately reproduce the correct ligand pose in the binding site for both proteins, we docked Geldanamycin to the glycosylated form of the protein to identify which region of the protein produces a binding site with the most favorable binding energies. We docked Geldanamycin to all five regions of the non-glycosylated and solvated minimized human Ecgp96 protein and ranked the complexes according to binding energy. The ligands docked to Region 1 (seen in **Figure 8.1**) have the best energy, indicating that this region is the most probable ligand-binding domain. The best binding energy of the complex in this region is -64.67 kcal/mol, and the best unified cavity energy is -53.87 kcal/mol. In comparison, the best binding complex in Region 2 is over 20 kcal/mol worse (binding energy at

-43.88 kcal/mol and unified cavity energy of -37.87 kcal/mol). The other regions did not produce a complex with better binding energies than that of Region 2.

Figure 8.8A shows the non-glycosylated complex with the best binding energy. Geldanamycin forms hydrogen bonds with five residues in the binding site including N142. Our predicted binding site agrees with our computationally docked and experimentally crystallized Hsp90-Geldanamycin complexes. Geldanamycin and ATP interact with this conserved asparagine residue in all crystallized Hsp90 systems.

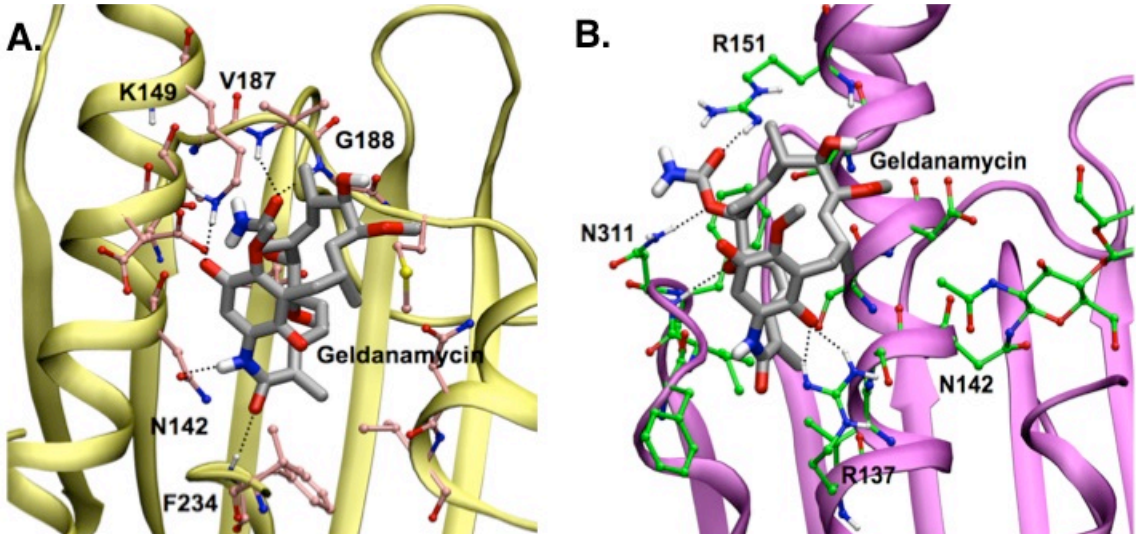


Figure 8.8 (**A)** Non-glycosylated Ecgp96-Geldanamycin complex with lowest binding energy. Geldanamycin (licrorice, grey) preferentially binds to N142 when the sugars are absent. Ecgp96's residue side chains are in CPK (ball-and-stick) form and are colored pink. (**B)** Glycosylated Ecgp96-Geldanamycin complex with lowest binding energy. Geldanamycin (licorice, grey) preferentially binds to Region 2 when N142 is glycosylated. Ecgp96's residue side chains are in CPK (ball-and-stick) form and are colored green. Hydrogen bonds are indicated by black dotted lines in the pictures.

Since Hsp90 has never been crystallized with its glycosylated segment (the biologically relevant form), we do not know the binding site of ATP or Geldanamycin in the presence of sugars. In order to predict an accurate binding site for Geldanamycin in Ecgp96, we docked the ligand to all five specified regions after adding sugars to N142 and N252. We found that the binding energy of Geldanamycin docked to Region 1 was worse after glycosylation because the binding energy increases from -64.67 kcal/mol to -53.60 kcal/mol, and the unified cavity increased from -53.87 kcal/mol to -48.44 kcal/mol, with the number of hydrogen bonds decreasing from five to two. The lowest energy pose (shown in **Figure 8.8B**) has a binding energy that 5 kcal/mol better than the best energy pose from Region 1 (binding energy of -58.90 kcal/mol; the unified cavity energy is 3.5 kcal/mol better at -52.07 kcal/mol). This site forms five hydrogen bonds with three

residues—R137, R151, and N311. Thus, our docking results suggest that Geldamycin preferentially would bind to a different site under glycosylated conditions.

DOCK Blaster 3 identifies small molecules that bind to unique region of the glycosylated protein

After identifying a novel region that can potentially bind known ligands after protein glycosylation, we decided to select new molecules via VLS that would preferentially bind to this part of the protein. Using the online server DOCK Blaster 3, we identified the portion of the glycosylated protein that we are interested in docking small molecules to. **Table 8.6** shows the binding energies calculated for the 500 selected ligands from DOCK Blaster 3. The binding energy of the 19 ligands covers a wide range, -21.47 to 21.07 kcal/mol. The number of hydrogen bonds varies from zero to three, and they seem to correlate with the binding energy. We do not observe a correlation between the VLS rank generated by DOCK Blaster 3 and our calculated binding energies, thus indicating the importance of enriching and evaluating the complexes with a different force field. For example, C03448152 (S7.3) has the lowest binding energy, but using the VLS standards it was ranked 18th out of the 19 ligands examined here.

Table 8.6 List of hits from DOCK Blaster 3 that we evaluated with the GenDock enrichment procedure and Dreiding-III FF and the comparison of our binding energy rank versus VLS rank.

Ligand name	Ligand name abbreviation	Binding Energy (kcal/mol)	Number of hydrogen bonds	Binding Energy Rank	VLS rank
C00176096	**S7.1**	-5.58	0	12	19
C03158777	**S7.2**	-16.82	3	4	4
C03448152	**S7.3**	-21.47	2	1	18
C09889887	**S7.4**	-12.00	0	7	11
C11697764	**S7.5**	10.51	1	18	7
C11820488	**S7.6**	-18.80	3	3	2
C12543557	**S7.7**	-13.44	2	6	8
C12577074	**S7.8**	-9.15	1	9	3
C12646952	**S7.9**	-15.65	0	5	9
C14950042	**S7.10**	-19.96	3	2	1
C17298946	**S7.11**	1.09	0	15	14
C22981496	**S7.12**	21.07	1	19	13
C24907001	**S7.13**	5.22	0	17	10
C30524473	**S7.14**	-6.67	0	10	12
C32777201	**S7.15**	-6.56	0	11	15
C45605944	**S7.16**	2.27	0	16	5
C47535390	**S7.17**	-10.80	0	8	16
C49028759	**S7.18**	-4.32	0	13	6
C58175257	**S7.19**	-3.23	1	14	17

The two ligands that we will investigate more closely are C034488152 (S7.3) and C11820488 (S7.6) (**Figure 8.9**). C034488152 (S7.3) had the lowest binding energy, as mentioned above, and C11820488 (S7.6) had the third lowest binding energy and is structurally very different. **Figure 8.10** shows how the respective ligands interact in the binding site. C034488152 (S7.3), in **Figure 8.10A**, has two hydrogen bonds with D148 and N311 and strong van der Waals interactions with I310 and F309. C11820488 (S7.6) in **Figure 8.10B** has three hydrogen bonds with D148, S141, and S144. The non-bonding interaction with D148 is extremely strong with -10.174 kcal/mol. This ligand also has significant hydrophobic interactions with I310, F309, and F312.

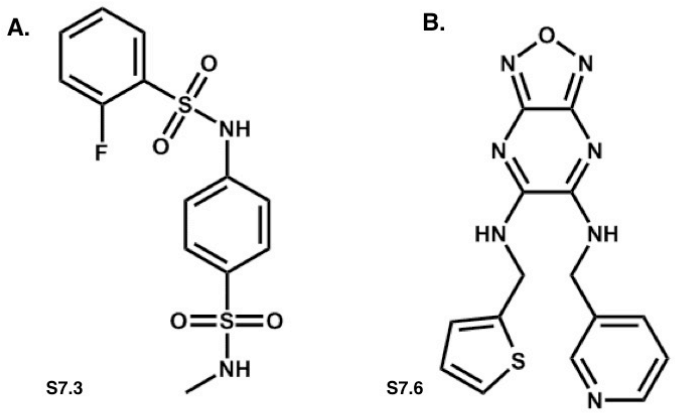
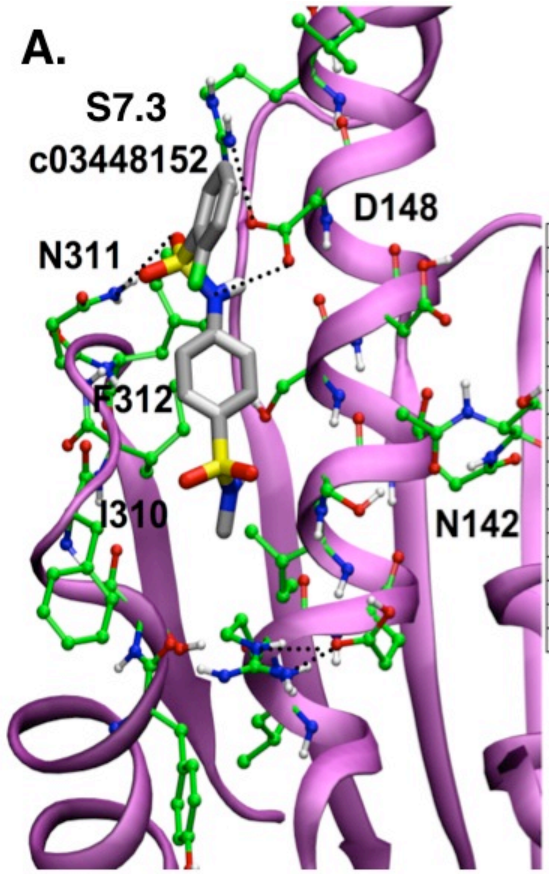
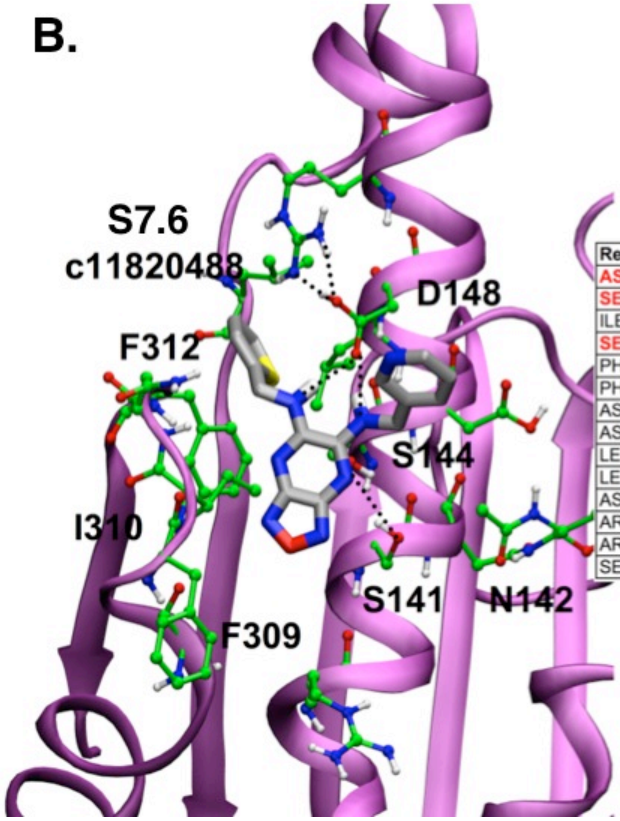


Figure 8.9. Chemical structures of two hit molecules from DOCK Blaster 3: **(A)** C03448152 (S7.3) and **(B)** C11820488 (S7.6).

While it is a positive sign that these low binding energy ligands form multiple strong hydrogen bonds with the protein, they have many polar atoms that are not being used effectively. Perhaps we could optimize the ligands further to maximize the interactions in that binding pocket. Also, we could dock various conformations of the ligands using our DarwinDock program to generate thousands of possible ligand positions for the given binding site rather than rely on the one provided by DOCK Blaster 3. We also need to test more ligands of the 500 provided. We have already performed a VLS for the non-glycosylated Ecgp96 protein, and we need to conduct the same analysis as described above.



Residue		VdW	Coulomb	H-Bond	NonBond
ASP	148	-2.016	-1.07	-3.558	-6.644
ILE	310	-4.391	0.252	0	-4.139
SER	141	-2.468	-0.385	0	-2.853
ASN	311	-1.706	0.577	-1.544	-2.672
PHE	309	-2.054	-0.243	0	-2.297
SER	144	-1.922	-0.006	0	-1.928
PHE	312	-1.759	0.13	0	-1.63
ARG	137	-1.452	-0.16	0	-1.612
ILE	140	-1.274	0.004	0	-1.271
ARG	151	-0.85	-0.357	0	-1.207
ASP	145	-0.701	0.063	0	-0.638
SER	307	-0.667	0.194	0	-0.473
LEU	136	-0.211	-0.254	0	-0.465
ASN	142	-0.503	0.045	0	-0.458
TYR	306	-0.303	-0.149	0	-0.452
GLU	138	-0.293	-0.086	0	-0.38
LEU	152	-0.236	0.104	0	-0.132

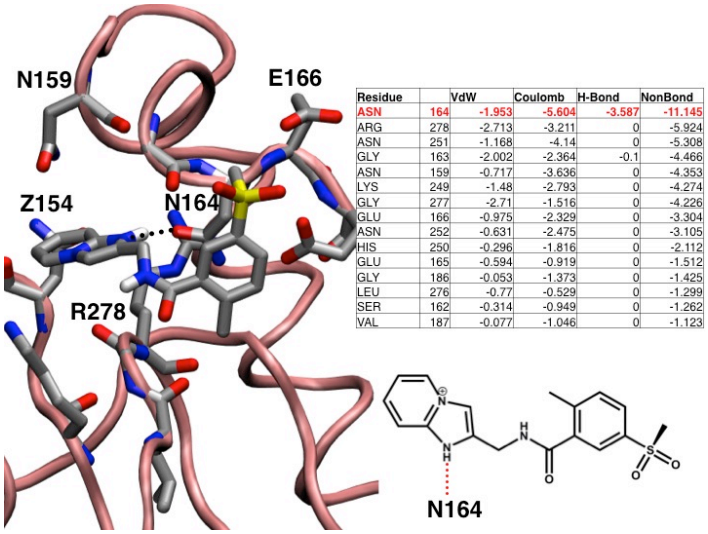


Residue		VdW	Coulomb	H-Bond	NonBond
ASP	148	-2.513	-2.374	-5.287	-10.174
SER	141	-1.648	-2.505	-2.739	-6.891
ILE	310	-3.902	0.072	0	-3.83
SER	144	-1.154	-1.428	-0.178	-2.76
PHE	309	-1.421	-0.792	0	-2.212
PHE	312	-2.016	0.11	0	-1.906
ASP	145	-1.436	0.07	0	-1.365
ASN	311	-2.45	1.131	0	-1.319
LEU	147	-1.264	0.257	0	-1.007
LEU	167	-1.054	0.068	0	-0.986
ASN	142	-1.033	0.391	0	-0.642
ARG	151	-0.4	0.248	0	-0.152
ARG	137	-0.419	0.294	0	-0.125
SER	307	-0.213	0.37	0	0.157

Figure 8.10. Hit molecules (**A**) C03448152 (S7.3) and (**B**) C11820488 (S7.6) in the glycosylated Ecgp96 model with cavity analysis. Hydrogen bonds are indicated by black dotted lines in the pictures and are in bold red in the tables. The cavity analysis breaks down the non-bonding energy into van der Waals (vdW), coulombic (Coulomb), and hydrogen bond (H-Bond) interactions. All residues with predicted energetic interactions with Ecgp96 are shown.

Z154 preferentially binds to a third region of Ecgp96

Z154 is a unique ligand in that it does not bind to Regions 1 or 2 as shown in **Figure 8.1**. Instead, it has the most favorable or lowest binding energies with the loop portion of the protein above Region 1. According to **Figure 8.11**, it only forms one hydrogen bond with N164, but since it is positively charged, it has strong electrostatic interactions with polar residues in the area. It has a very favorable binding energy of -63.46 kcal/mol, which is a significant 6 kcal/mol lower in energy than the known binding ligand Geldanamycin (**Table 8.7**). The predicted binding site of Z154 suggests another unexplored region for potential small molecule binding. We plan to conduct VLS for this region of the protein to identify possible hits that can be tested by experimental collaborators.



Residue		VdW	Coulomb	H-Bond	NonBond
ASN	164	-1.953	-5.604	-3.587	-11.145
ARG	278	-2.713	-3.211	0	-5.924
ASN	251	-1.168	-4.14	0	-5.308
GLY	163	-2.002	-2.364	-0.1	-4.466
ASN	159	-0.717	-3.636	0	-4.353
LYS	249	-1.48	-2.793	0	-4.274
GLY	277	-2.71	-1.516	0	-4.226
GLU	166	-0.975	-2.329	0	-3.304
ASN	252	-0.631	-2.475	0	-3.105
HIS	250	-0.296	-1.816	0	-2.112
GLU	165	-0.594	-0.919	0	-1.512
GLY	186	-0.053	-1.373	0	-1.425
LEU	276	-0.77	-0.529	0	-1.299
SER	162	-0.314	-0.949	0	-1.262
VAL	187	-0.077	-1.046	0	-1.123

Figure 8.11. Hydrogen bonds are indicated by black dotted lines in the pictures and are in bold red in the tables. The cavity analysis breaks down the non-bonding energy into van der Waals (vdW), coulombic (Coulomb), and hydrogen bond (H-Bond) interactions. Only residues whose energetic interactions are greater than 1 kcal/mol are shown in the table.

Table 8.7. Comparison of binding energies of Z154 and Geldenamycin to the glycosylated form of Ecgp96.

Protein (Ecgp96)	Ligand	Binding Energy (kcal/mol)
Glycosylated	Geldanamycin	-57.47
Glycosylated	Z154	-63.46

CONCLUSIONS

We docked three known ligands that are effective at preventing bacterial invasion to our predicted structure of Ecgp96 and, based on the predicted binding site, we recommend novel ligands that we anticipate to have stronger binding affinities for this protein than the ligands that are already known. Even more excitingly, we offer strong computational evidence that Telmisartan can bind to Ecgp96, not just AT1R. We also docked Geldanamycin to Ecgp96 in both its non-glycosylated and glycosylated forms only to find that the ligand has more favorable energies with a different part of the protein upon glycosylation. This computational finding needs to be explored further through additional experiments, which will be done at a future time. We performed VLS using DOCK Blaster 3 to identify potential small molecules that would bind to this novel and unexplored region. We identify several candidates, which will be tested using binding assays. If these ligands bind strongly, then perhaps we have located a new binding site on the protein that has never been considered or used until now and that would be effective in preventing interaction with the OmpA protein of *E. coli* and subsequent onset of meningitis.

REFERENCES

1. Unhanand M, Mustafa MM, McCracken GH, Nelson JD (1993) Gram-negative bacillary meningitis: A 21-year experience. J Pediatr **122**:15-21.
2. Gladstone IM, Ehrenkranz RA, Edberg SC, Baltimore RS (1990) A 10-year review of neonatal sepsis and comparison with the previous 50-year experience. Pediatr Infect Dis J **9**:819-825.
3. Koedel U, Pfister HW (1999) Models of experimental bacterial meningitis: Roles and limitations. Infect Dis Clin North Am **13**:549-577.
4. Leib SL, Tauber MG (1999) Pathogenesis of bacterial meningitis. Infect Dis Clin North Am **13**:428-527.
5. Prasadarao NV (2002) Identification of *Escherichia coli* outer membrane protein A receptor of human brain microvascular endothelial cells. Infect Immun **70**:4556-4563.
6. Prasadarao NV, Srivastava PK, Rudrabhatla RS, Kim KS, Huang S, Sukumaran SK (2003) Cloning and expression of the *Escherichia coli* K1 outer membrane protein A receptor, a gp96 homologue. Infect Immun **71**:1680-1688.
7. Pascal TA, Abrol R, Mittal R, Wang Y, Prasadarao NV, Goddard WA, III (2010) Experimental validation of predicted binding site of *Escherichia coli* K1 outer membrane protein A to human brain microvascular endothelial cells. J Biol Chem **285**:37753-37761.
8. Datta D, Vaidehi N, Floriano WB, Kim KS, Prasadarao NV, Goddard WA, III (2003) Interactions of *E. coli* outer-membrane protein A with sugars on the receptors of the brain microvascular endothelial cells. Proteins **50**:213-221.
9. Krishnan S, Shanmuganathan MV, Behenna D, Stoltz BM, Prasadarao NV (2014) Angiotensin II receptor type 1-A novel target for preventing neonatal meningitis in mice by *Escherichia coli* K1. J Infect Dis **209**:409-419.

10. Sugawara T, Kokubun K, Ishida R, Fujiwara K, Yamamoto S, Kinouchi H, Itoh H, Mizoi K. Angiotensin receptor blocker directly binds to HSP and stimulates its production in the brain. In: Munteanu V, Raducanu R, Dutica G, Croitoru A, Balas VE, Eds. (2010) Recent Advances in Mathematics and Computers in Business, Economics, Biology & Chemistry. WEAS Press, Iasi, Romania, pp. 350-357.
11. Stebbins CE, Russo AA, Schneider C, Rosen N, Hartl FU, Pavletich NP (1997) Crystal structure of an Hsp90, geldanamycin complex: Targeting of a protein chaperone by an antitumor agent. Cell **89**:239-250.
12. Ghosh A, Rapp CS, Friesner RA (1998) Generalized born model based on a surface integral formulation. J Phys Chem B **102**:10983-10990.
13. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem **4**:187-217.
14. Shaaltiel Y, Bartfeld D, Hashmueli S, Baum G, Brill-Almon E, Galili G, Dym O, Boldin-Adamsky SA, Silman I, Sussman JL, Futerman AH, Aviezer D (2007) Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. Plant Biotechnol J **5**:579-590.
15. Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH (1997) Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. Cell **90**:65-75.
16. Maestro, version 7.5. (2006). Schrodinger, LLC, New York, NY.
17. Jaguar, version 6.5. (2006). Schrodinger, LLC, New York, NY.
18. MacroModel, 9.5. (2007). Schrodinger, LLC, New York, NY.
19. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still WC (1990) Macromodel -- An integrated software system for modeling organic and bioorganic molecules using molecular mechanics. J Comput Chem **11**:440-467.
20. Lim KT, Brunett S, Iotov M, McClurg RB, Vaidehi N, Dasgupta S, Taylor S, Goddard WA, III (1997) Molecular dynamics for very large systems on massively parallel computers: The MPSim program. J Comput Chem **18**:501-521.
21. Mayo SL, Olafson BD, Goddard WA, III (1990) DREIDING: A generic force field for molecular simulations. J Phys Chem **94**:8897 - 8909.
22. Floriano WB, Vaidehi N, Zamanakos G, Goddard WA, III (2004) HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. J Med Chem **47**:56-71.
23. Goddard WA, III, Kim S-K, Li Y, Trzaskowski B, Griffith AR, Abrol R (2010) Predicted 3D structures for adenosine receptors bound to ligands: Comparison to the crystal structure. J Struct Biol **170**:10-20.
24. Kim S-K, Li Y, Abrol R, Heo J, Goddard WA, III (2011) Predicted structures and dynamics for agonists and antagonists bound to serotonin 5-HT2B and 5-HT2C receptors. J Chem Inf Model **51**:420-433.
25. Kim S-K, Riley L, Abrol R, Jacobson KA, Goddard WA, III (2011) Predicted structures of agonist and antagonist bound complexes of adenosine A(3) receptor. Proteins: Struct Funct Bioinform **79**:1878-1897.
26. Nair N, Kudo W, Smith MA, Abrol R, Goddard WA, III, Reddy VP (2011) Novel purine-based fluoroaryl-1,2,3-triazoles as neuroprotecting agents: Synthesis, neuronal cell culture investigations, and CDK5 docking studies (vol 11, pg 3165, 2001). Bioorg Med Chem Lett **21**:5649-5649.
27. Lang PT, Brozell SR, Mukherjee S, Pettersen ET, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID (2009) DOCK 6: Combining techniques to model RNA-small molecule complexes. RNA **15**:1219-1230.

28. Kam VWT, Goddard WA, III (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. J Chem Theory Comput **4**:2160-2169.
29. Bolton E, Wang Y, Thiessen P, Bryant S. PubChem: Integrated platform of small molecules and biological activities. (2008) Annual Reports in Computational Chemistry. Elsevier Ltd., Oxford, UK., pp. 217-240.
30. Tetko IV, Tanchuk VY (2002) Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. J Chem Inf Comput Sci **42**:1136-1145.
31. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc **118**:11225-11236.
32. Irwin JJ, Shoichet BK, Mysinger MM, Huang N, Colizzi F, Wassam P, Cao Y (2009) Automated docking screens: a feasibility study. J Med Chem **52**:5712-5720.

Chapter IX

THERMOCHEMISTRY OF *TRANS*-DIOXOVANADIUM COMPOUNDS

This project was done in collaboration with Dr. Thomas P. Umile and Professor John T. Groves of the Department of Chemistry at Princeton University and Dr. Mu-Jeng Cheng at Caltech.

ABSTRACT

With the decrease in the availability in fossil fuels and the increase in the amount of carbon dioxide released into the atmosphere, it has become necessary to maximize the use of alternative fuels. Natural gas, made mostly of methane, is an attractive option due to its abundance in the US and the decreased amount of carbon dioxide emissions it produces. However, methane is expensive to retrieve and transport in its gaseous state and difficult to modify without extreme conditions due to its very high bond dissociation energy of 105 kcal/mol. Thus, we have proposed and virtually explored the reactivity of high valent metal-oxo catalysts (*trans*-dioxovanadium(V) porphyrin and corrin complexes) that are reactive enough to break alkane bonds. Using quantum mechanical computations, we surveyed the free energy surface of states involved in hydrocarbon bond activation. We learn that the proposed catalyst, due to its high nucleophilicity, has an O-H bond dissociation energy comparable to that of methane and furthermore that this catalyst can produce alkyl radicals that escape to form chemical species other than the typical alcohol products of radical rebound. Hence, we provide sufficient evidence that suggests *trans*-dioxovanadium(V) is reactive enough to be an attractive catalyst that can be useful in functionalizing alkanes and warrants characterization and testing in the future.

INTRODUCTION

One of the greatest scientific challenges is finding a clean and sustainable fuel that can power modern technology and act as an alternative to fuels such as petroleum and coal. Methane, the major component of natural gas, is an inexpensive and plentiful fuel that does not release as much carbon dioxide into the atmosphere. Upon consumption, methane produces 47% and 27% less carbon dioxide gas than coal and petroleum, respectively (1). There are enormous methane reserves in the United States, but unfortunately, they are in remote locations, making methane transportation very expensive using infrastructure designed for liquid fuels. One solution to the transportation issue is to convert methane into liquid methanol, but this process is very difficult because methane is kinetically stable. Its C-H bond dissociation enthalpy (BDE) is 105 kcal/mol (2). Our goal is to design a catalyst that can convert methane into methanol at a reasonable rate.

The reaction examined here involves an alkane reacting with an oxo-metal compound to produce a hydroxo-metal complex and an alkyl radical that is free to interact with other species and form new products. This type of reaction has been studied previously (3; 4). For example, alkanes react with *trans*-oxofluoromanganese(V) complexes to produce an alkyl radical, which can

further react with the Mn(IV)(TMP)F_2 to produce a compound with C-F bonds (3). Secondly, alkanes react with *trans*-dioxomanganese to produce an alkyl radical and an L-Mn(IV)-OH complex. Subsequently, an L-Mn(IV)-OCl complex may react with the alkyl radical to form a product with C-Cl bonds (4). Here, the first half of the reaction between *trans*-dioxovanadium compounds and alkanes, which produces alkyl radicals, is examined. These radicals are expected to interact further to form new products in a manner similar to the reactions in the studies mentioned above.

The mechanism discussed in **Figure 9.1** is the high valent metal-oxo hydrocarbon bond activation catalytic cycle. During this process, the metal (M) undergoes oxidation to create a metal-oxo to activate the alkane's (R-H) C-H bond. The hydrogen is removed from the hydrocarbon and attaches to the oxygen of the metal-oxo complex. If the metal is electrophilic, like manganese, then the OH group will detach from the metal and rejoin with the alkyl radical to produce the rebound product (5). However, if the metal is nucleophilic, like vanadium, then the hydrogen could remain on the catalyst and the alkyl radical groups could react with one another or with a different chemical species leading to the radical escape products rather than alcohol products. Another process is necessary to regenerate the initial starting materials to complete the catalytic cycle. Our collaborators have previously isolated and characterized stable high valent metal-oxos including a manganese porphyrin compound Mn(TMP)Cl that catalyzes oxidative aliphatic fluorination (3). However, as noted above, manganese is an electrophilic metal, so the number and variety of potential products catalyzed by such a complex is limited. In this report, we use quantum mechanical simulations to explore the potential of the nucleophilic vanadium porphyrin complexes to be catalysts that are strong enough to activate alkane bonds as well as produce radical escape compounds. Furthermore, we propose a regeneration mechanism, compare these results with those of manganese complexes, and analyze the effect that the ligand plays on the thermochemistry. We provide sufficient evidence that *trans*-dioxovanadium(V) compounds could be effective catalysts that can react with methane and produce an array of new chemical products.

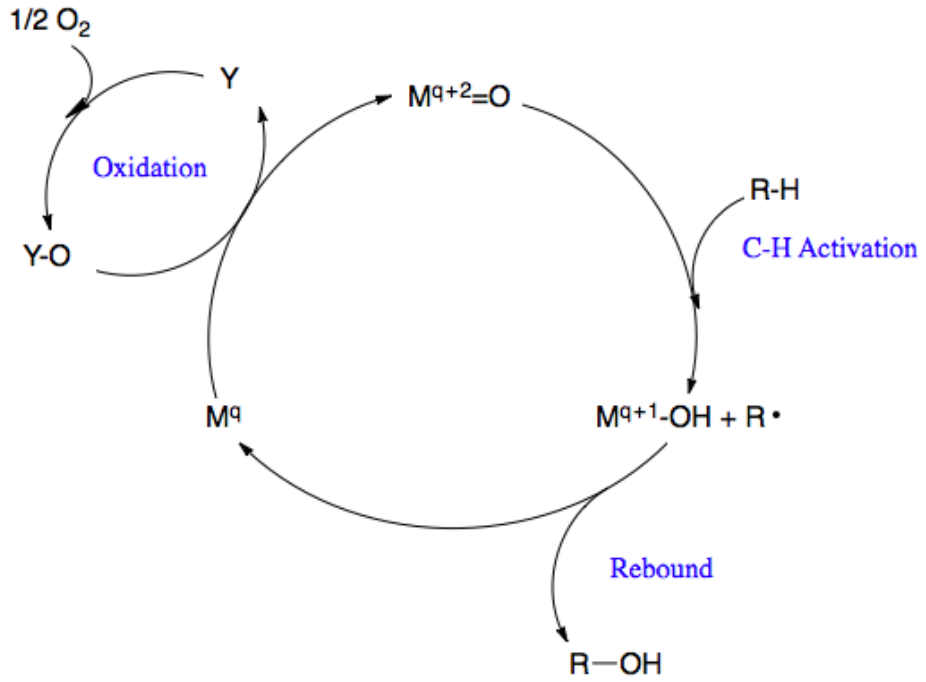


Figure 9.1 Diagram of the high valent metal-oxo hydrocarbon bond activation catalytic cycle. *M* stands for the metal, *R-H* represents the alkane, *q* is the oxidation state of the metal. Oxygen and hydrogen are represented by their respective chemical symbols. Modified version of **Figure 1** from Ref. (5).

METHODS

We performed geometry optimizations using Jaguar 7.9 (6) with the B3LYP-D3 density functional (7-10) and the 6-31G** basis set (11-14) except for with the metal atoms that use the LACV3P++ basis set (15). Using the optimized molecules, we calculated the frequencies at 298.15 K and performed a single energy point calculation including aqueous solvation with the M06 functional (16) and the 6-311G**++ basis set (11-14) except for with the metal atoms, vanadium and manganese, which use the LACV3P**++ basis set (15). We calculated the solvation energies with the Poisson-Boltzmann solvation model (17; 18) with a dielectric constant of 80.37 and probe radius of 1.40 Å. To determine the stability of all the components of the reaction, including the intermediates and transition states, we determined the Gibbs free energy (*G*) and enthalpy (*H*) using **Equations 9.1** and **9.2** respectively:

$$G = E + PV - TS$$
$$= E_{elec} + G_{solv} + 6kT + H_{vib} + ZPE - TS_{tot}, \qquad \textbf{(9.1)}$$

$$H = E_{elec} + G_{solv} + 6kT + H_{vib} + ZPE \quad \textbf{(9.2)}$$

To obtain the Gibbs energy **(9.1)**, the electronic energy, (E_{elec}) and solvation free energy (G_{solv}) from the single point energy calculation are added to the vibrational enthalpy (H_{vib}), the zero-point energy (*ZPE*), and total entropy (S_{tot}) from the frequency calculation, and *6kT*, which accounts for ½ *kT* of potential and kinetic energy of the six external degrees of freedom. We also use these quantities in **(9.1)** to compute bond dissociation free energies (BDFE) and bond dissociation enthalpies (BDE) **(9.2)**.

The *G* determined from **(9.1)** is used to find the protonation driving force, or pK_a in **Equation 9.3**:

$$pK_a = \frac{\Delta G}{(2.303RT)} \quad \textbf{(9.3)}$$

with *R* equal to the ideal gas constant of $1.987*10^{-3}$ kcal/Kmol and *T* equal to 298.15 K.

Oxidation driving potentials, *E*, are given relative to SHE and depend on *G* from **(9.1)** as seen in **Equation 9.4**:

$$E = \frac{\Delta G}{nF} - E_{SHE} \quad \textbf{(9.4)}$$

with *F* equal to Faraday's constant, or 23.06 kcal/(V*mol), *n* equal to the number of electrons, and E_{SHE} equal to the standard hydrogen reference potential, an empirical value of 4.28 V (19).

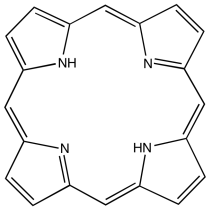
RESULTS AND DISCUSSION

Thermochemistry of trans-dioxovanadium(V) compounds

Figure 9.2 shows the pH-dependent redox thermochemistry of the *trans*-dioxovanadium(V) tetrahydroporphyrin (THP) molecule and the oxohydroxovanadium(IV) porphyrin. Equivalent to hydrogen atom transfer from an alkane, *trans*-dioxovanadium(V) can undergo reduction followed by protonation. A reduction potential of 0.0 V is required to generate the *trans*-dioxovanadium(IV) molecule. (The free energy of this linear $O\text{-}V^{V}\text{-}O$ isomer is more favorable than that of its isomer, *trans*-dioxovanadium(V) coordinated to a radical anionic porphyrin ring,

by 25 kcal/mol.) *Trans*-dioxovanadium(IV) has a large driving force for protonation, with a pK_a of 35.0, to create the *trans*-oxohydroxovanadium(IV). Alternatively, *trans*-dioxovanadium(V) can undergo protonation then reduction. This route has a slightly smaller driving force for protonation with a pK_a of 21.0 followed by a higher reduction potential of 0.9 V. Protonation is accompanied by an electronic rearrangement in which an electron from the porphyrin π-system moves to the vanadium d_{xy} orbital, generating a V(IV) center. Overall, *trans*-dioxovanadium(V) has a O-H bond BDE of 103 kcal/mol, a BDFE of 97 kcal/mol, and a high pK_a of 21.0. These energies make *trans*-dioxovanadium(V) an appealing catalyst because its BDE is approximately equal to that of the bond energy of methane, 105 kcal/mol. They also highlight the challenges to developing working catalytic cycles by limiting the media in which the intermediates would persist and the means of oxidatively regenerating the reactive *trans*-dioxovanadium(V).

A.



B.

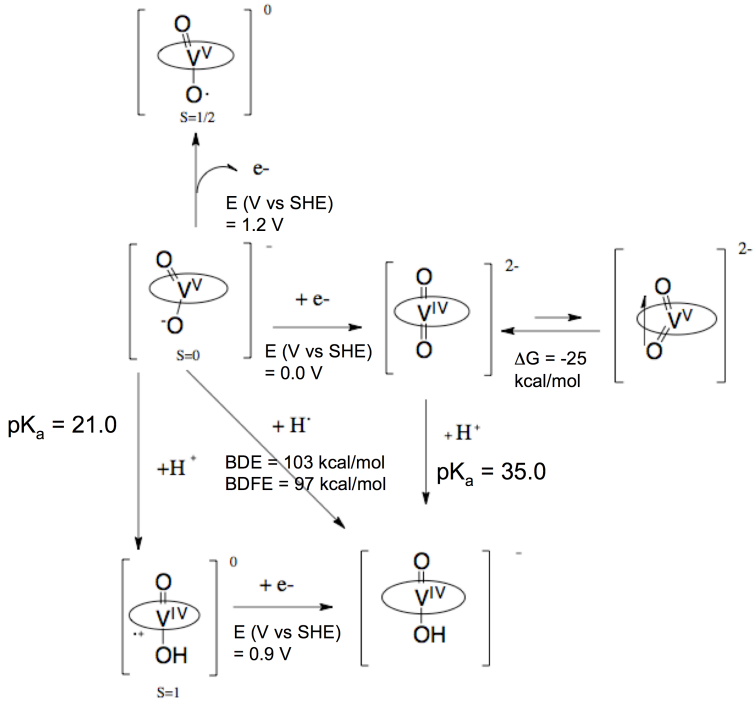


Figure 9.2 (A) Structure of the porphyrin macrocycle. **(B)** Thermochemistry of *trans*-dioxovanadium(V) THP compounds. The oxidation potentials (E, V vs. SHE), free energy (*G*, kcal/mol), pK_a, total spins (S), bond dissociation enthalpy (BDE), and bond dissociation free energy (BDFE) required to generate oxohydroxovanadium(IV) from *trans*-dioxovanadium(V) are shown above.

Hinting at the electron richness that drives the vanadium(V) reactivity, oxidation of the *trans*-dioxovanadium(V) anion to generate the neutral doublet requires a moderate potential of 1.2 V versus SHE. The *trans*-dioxovanadium(V) anion singlet is more favorable in energy than the triplet by 3 kcal/mol. The singlet has four σ, four π_x, and four π_y electrons, but no δ electrons. The result of one-electron oxidation is the neutral *trans*-dioxovanadium(V) doublet, which is 6 kcal/mol lower in energy than the quartet. In the doublet, the V-O single bond (1.84 Å) is longer than the double bond (1.61 Å), and the singly bonded oxygen hosts 0.95 e- of spin density (by Mulliken analysis) while the vanadium atom has a somewhat down spin. The product of the *trans*-dioxovanadium(V) anion protonation is the triplet of the neutral oxohydroxovanadium(IV), which is 4 kcal/mol more stable than the singlet. In the triplet, the V-O-H bond is significantly longer than the V-O double bond, 2.03 Å and 1.62 Å respectively. The d_{xy} orbital of the vanadium metal is singly occupied, and there is a hole on the porphyrin ligand.

Figure 9.3 illustrates the importance of the *trans*-dioxo configuration. *Cis* isomers of the V(V) dioxos are more stable than their *trans* counterparts since all five *d*-orbitals can be involved in covalent bonding to the oxygen atoms. For the dioxovanadium(V) complex, the *cis* isomer is energetically favorable by 15.8 kcal/mol. For the oxohydroxovanadium(IV) complex, on the other hand, the trans isomer is lower in energy by 21.7 kcal/mol. Only in the *trans* configuration can the lone electron of V(IV) be stored in a *d*-orbital orthogonal to all the σ and π bonds to the oxygen atoms. As a result, the O-H BDE to the *cis* isomer is 67 kcal/mol, not comparable to the bond energy of methane. The *trans* geometry is necessary to generate oxovanadium(IV) with a high energy O-H bond.

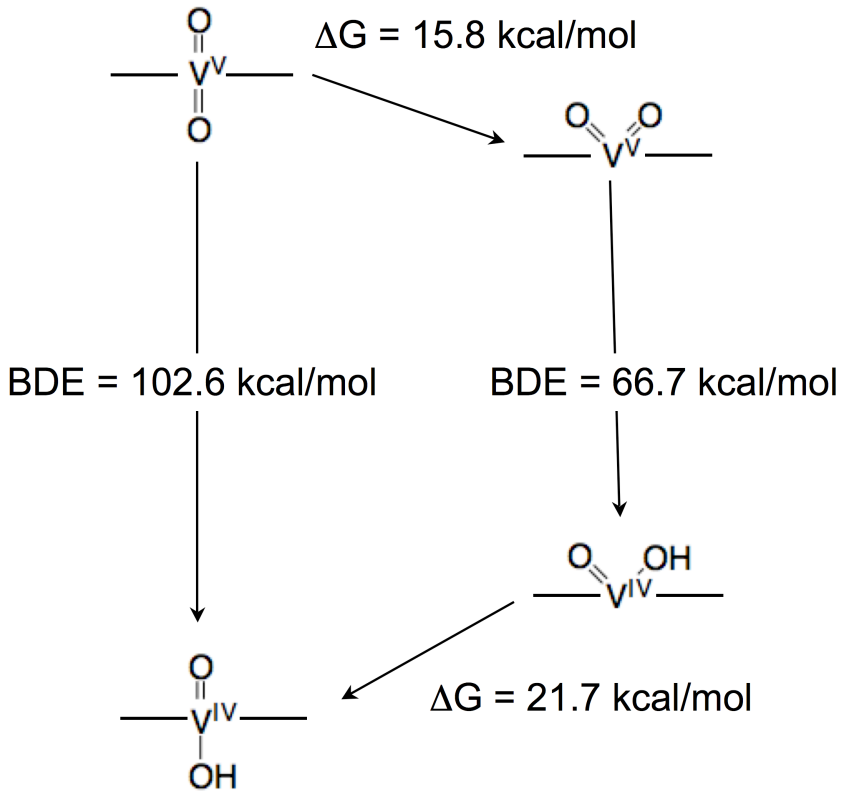


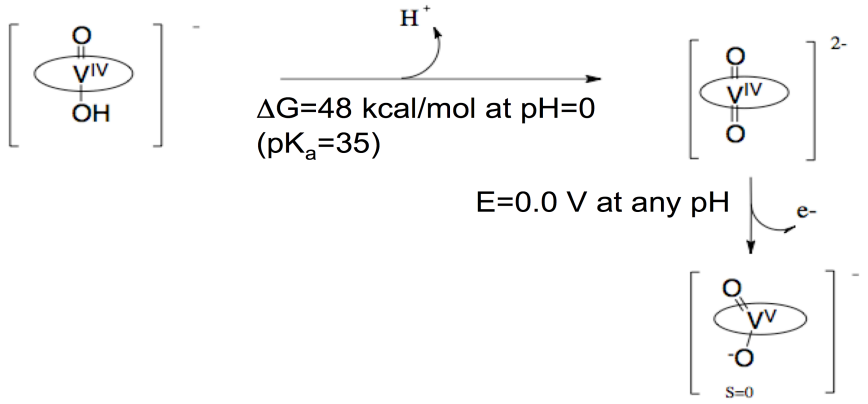
Figure 9.3 Thermochemical comparison of the *trans* and *cis* geometries of dioxovanadium(V) and oxohydroxovanadium(IV) compounds. The BDE between the dioxovanadium(V) and oxohydroxovanadium(IV) compounds and the change in free energy (ΔG) between the *cis* and *trans* geometries are indicated.

Thermochemistry to regenerate trans-dioxovanadium(V) compounds

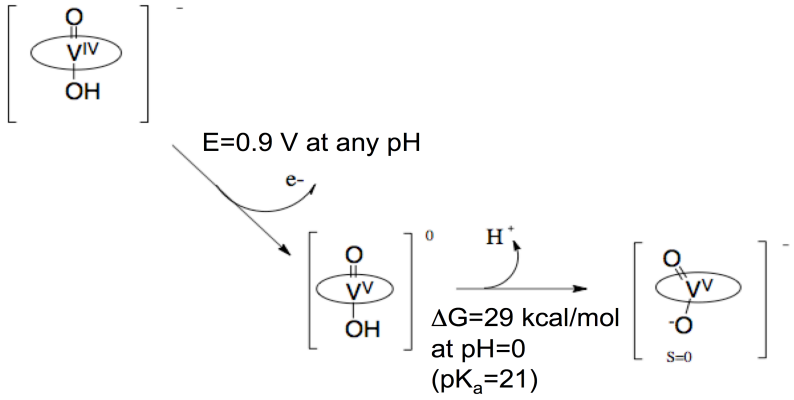
Our simulations suggest that *trans*-dioxovanadium(V) would be an attractive catalyst in converting methane to methanol due to its high BDE. We have examined the thermodynamics for three possible pathways, shown in **Figure 9.4**, for regenerating the high oxidation state. In **Figure 9.4A**, *trans*-oxohydroxovanadium(IV) undergoes deprotonation followed by oxidation. The deprotonation step requires a large pH of 35. The oxidation step should occur easily since the vanadium(IV) dianion is oxidized at a mild potential of 0.0 V. In the second pathway, **Figure 9.4B**, *trans*-oxohydroxovanadium(IV) undergoes one electron oxidation followed by deprotonation. Here, the oxidation step will not occur as readily because the potential is more moderate at 0.9 V, but chemical and photochemical oxidants cover this potential range. The pH required for the deprotonation step is 21, a significant step from 35 toward a manageable pH. Since O-atom transfer reagents have been used to turn over oxo-based catalysts, the third pathway, **Figure 9.4C**, consists of an oxygen atom transfer (a two-electron oxidation) followed by a one electron reduction. The oxidation potential of the first step is very high at 1.4 V at

pH=14 and 1.8 V at pH=0, thus, a strong oxidant is required. The resulting intermediate is reactive because it has a reduction potential of 1.5 V. Of these three pathways, the second one seems most promising. The first pathway requires a very large pH of 35, which would make deprotonation difficult. The circuitous third pathway requires a chemical oxidant and electrochemical reduction making the second pathway the most feasible target.

A.



B.



C.

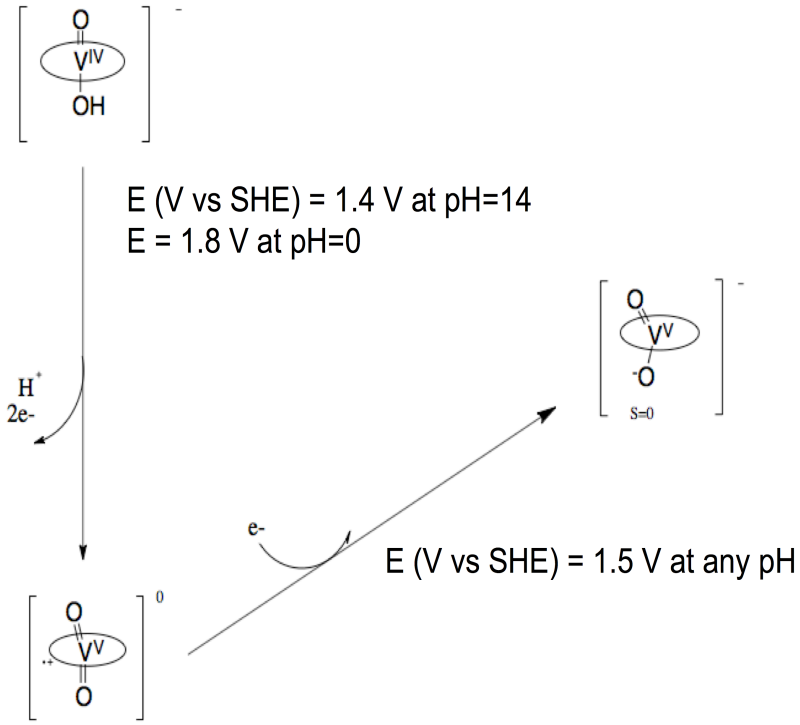


Figure 9.4 Three pathways for regenerating *trans*-dioxovanadium(V) from *trans*-oxohydroxovanadium(IV). **(A)** One electron oxidation followed by deprotonation. **(B)** Deprotonation followed by one electron oxidation. **(C)** Oxygen atom transfer followed by one electron reduction. Reduction and oxidation potentials (E) and pH for deprotonation are indicated.

Comparison with trans-oxoaquamanganese(V) complex's thermochemistry

Oxomanganese porphyrin complexes competent for activating strong C-H bonds have previously been characterized (3; 20), so here we contrast the thermochemistry of *trans*-oxoaquamanganese(V) and that of the *trans*-dioxovanadium(V). **Figure 9.5** shows the stepwise addition of a hydrogen atom to form *trans*-hydroxoaquamanganese(V). The driving force for hydrogen atom transfer is dramatically shifted to reduction of the oxo in the manganese case (1.2 V vs. SHE for Mn(V) vs. 0.0 V for V(V)) from protonation of the reduced oxo in the vanadium case (pK_a of 35 for V(IV) oxo vs. pK_a of only 7.3 for Mn(IV) oxo). The O-H bond dissociation energy of 93 kcal/mol for the manganese oxo is derived from its electrophilicity. The 103 kcal/mol BDE predicted for the putative vanadium oxo is derived from its nucleophilicity, with more driving force for protonating the vanadyl oxo.

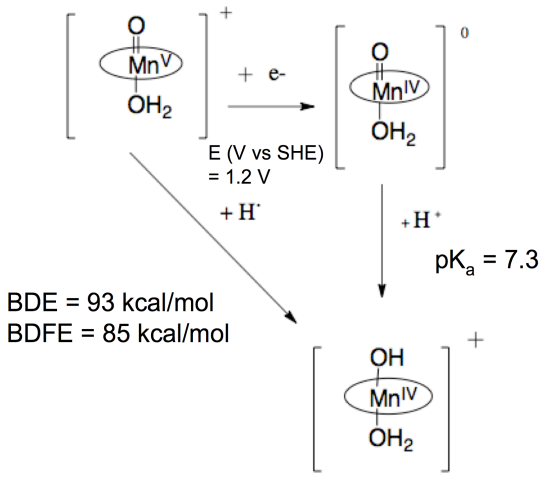
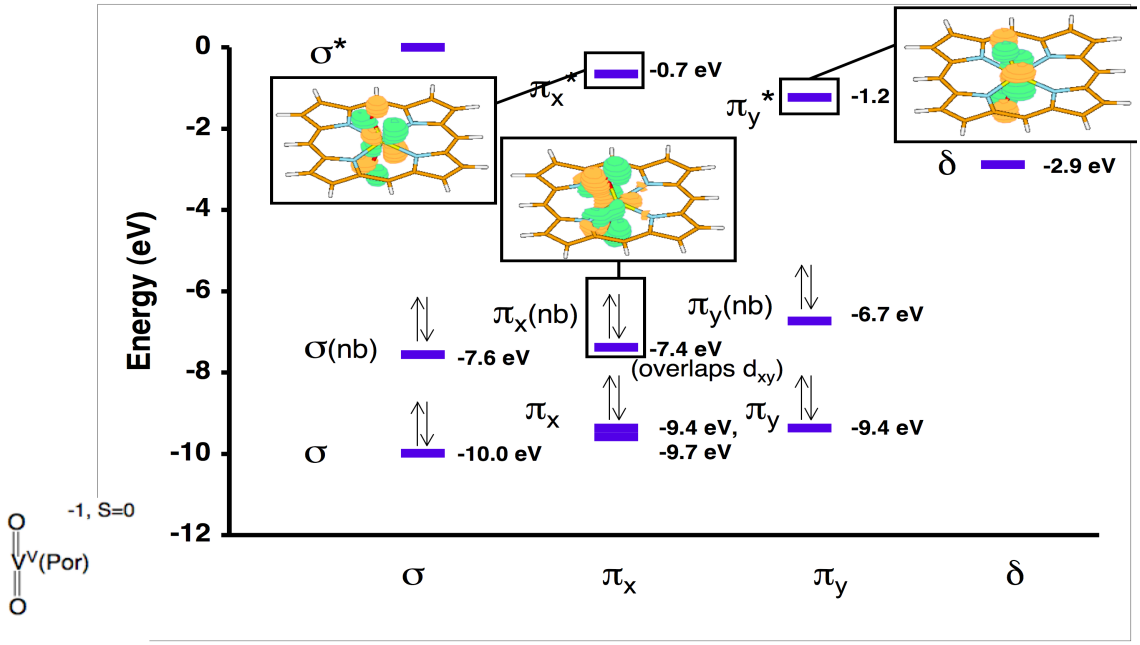


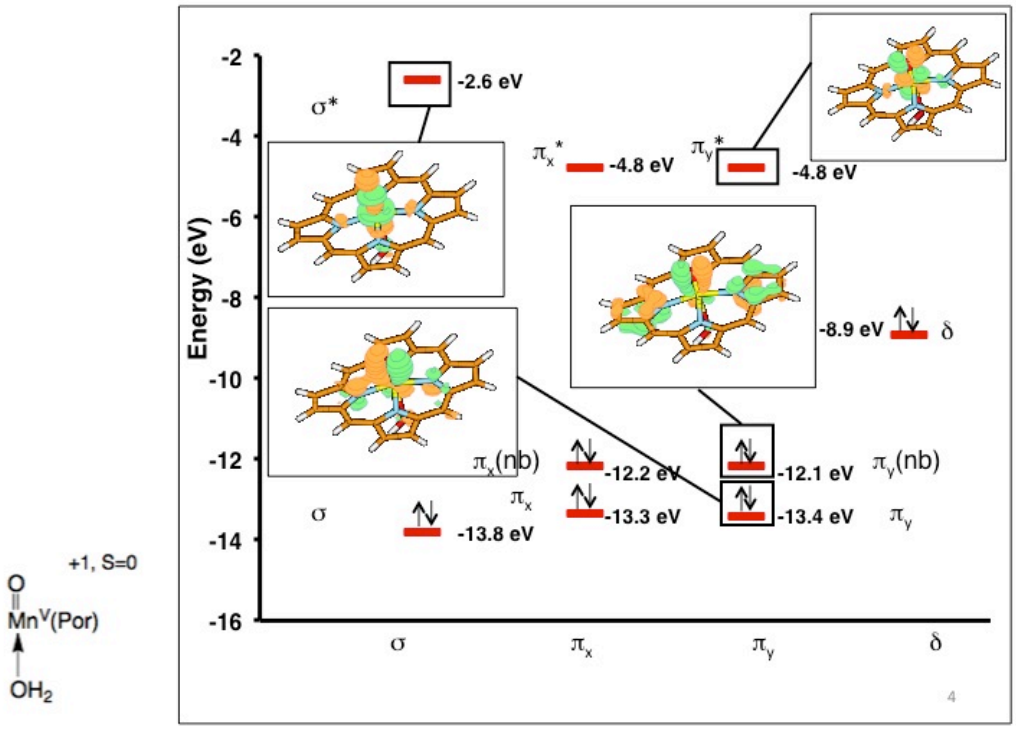
Figure 9.5 Thermochemistry of *trans*-oxoaquamanganese(V) compounds. The reduction potentials (E), protonation driving force (pK_a), bond dissociation enthalpy (BDE), and bond dissociation free energy (BDFE) required to generate *trans*-hydoxoaquamanganese(IV) from *trans*-oxoaquamanganese(V) are shown above.

The two complexes' different driving forces for hydrocarbon bond activation can also be illustrated with molecular orbital diagrams shown in **Figure 9.6**. The diagram in **Figure 9.6A** shows the energies for the σ, π_x, π_y, and δ orbitals of *trans*-dioxovanadium(V). The d_{xy} orbital at -2.9 eV is empty and can therefore accept electron density from the oxygen p_x lone pairs in the π_x non-bonding orbital after the oxos 'lean over' the *xz* plane. Thus, the *trans*-dioxovanadium(V) has a bent geometry (O-V-O = 146.2º) as shown in **Figures 9.2, 9.4, and 9.6A**. The inset shows the overlap between the lone pairs of the oxos orbitals and vanadium's empty d_{xy} orbital. For the *trans*-oxoaquamanganese(V) compound, on the other hand, the δ orbital at -8.9 eV is occupied by additional electrons as shown in **Figure 9.6B**. Although this compound contains one π_y and one π_x Mn-O π-bond, we found each of these formed bonding and anti-bonding combinations with the overlapping nitrogen-based orbitals. **Figure 9.6C** shows the comparison of the energy levels of *trans*-dioxovanadium(V) in blue versus *trans*-oxoaquamanganese(V) in red. Orbitals of the electrophilic manganese complex are lower than those of the corresponding vanadium complex by roughly 5 eV. Since the d_{xy} orbital is occupied in *trans*-oxoaquamanganese(V), the π_y* orbital is occupied upon H-atom transfer, whereas in *trans*-dioxovanadium(V), the d_{xy} orbital is empty, so hydrogen's electron would go into the δ orbital, which is 2 eV higher in energy than the Mn π_y*.

A.


Energy (eV)
0
-2
-4
-6
-8
-10
-12
σ*
πx*
-0.7 eV
πy*
-1.2
δ
-2.9 eV
πx(nb)
-7.4 eV
(overlaps dxy)
πy(nb)
-6.7 eV
σ(nb)
-7.6 eV
πx
-9.4 eV,
-9.7 eV
πy
-9.4 eV
σ
-10.0 eV
σ
πx
πy
δ
-1, S=0
O
V^V(Por)
O

B.



C.

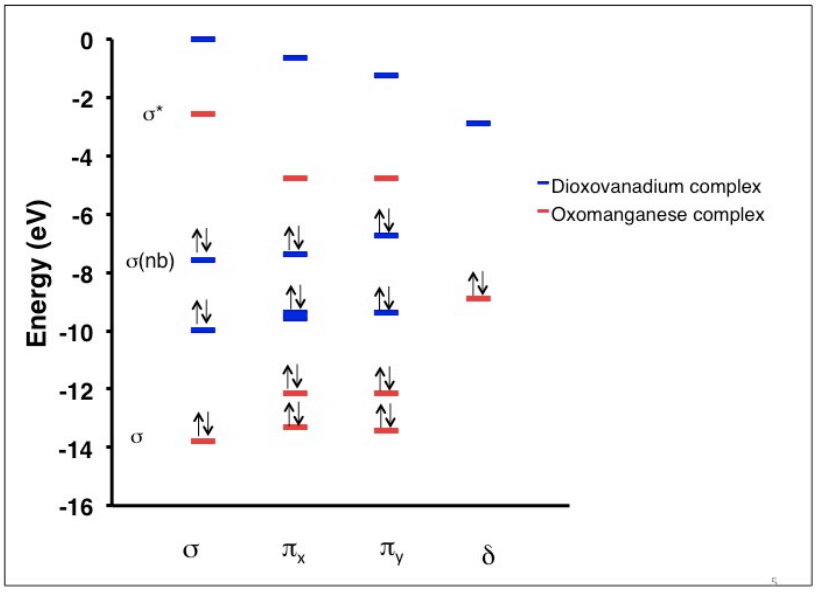
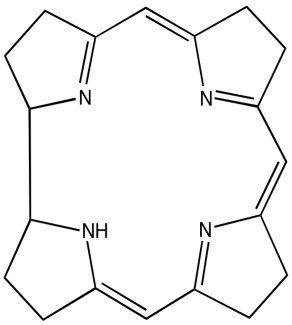


Figure 9.6 Molecular orbital (MO) diagrams of the manganese and vanadium compounds. (**A)** MO diagram of *trans*-dioxovanadium(V). The bent oxos lay on the *x*-axis. (**B)** MO diagram of *trans*-oxoaquamanganese(V). Mn-N lie on the *x*-axis. (**C**) MO diagram comparing the energies of the orbitals shown above in **Figure 9.6A** and **9.6B**.

Changes in thermochemistry resulting from the replacement of the porphyrin ligand

Since the proposed *trans*-dioxovanadium is predicted to be excessively reactive and is yet unsynthesized, less reactive variations were sought that may be more likely to be isolated. Ideally, the vanadium's d_{xy} orbital would lie higher in energy in order to stabilize the higher oxidation state. We hypothesized that replacing the porphyrin macrocycle with a corrin ligand whose nitrogen atoms are closer to the metal would destabilize the d_{xy} orbital. As anticipated, the corrin macrocycle (**Figure 9.7A**) lowers the $V^{IV/V}$ oxidation potential by 0.5 V (**Figure 9.7B**), and destabilizes the vanadium d_{xy} orbital by increasing its energy from -2.9 eV to -2.7 eV. (The *trans*-dioxovanadium(IV) isomer is 13 kcal/mol more stable than the *trans*-dioxovanadium(V) bound by a radical anion.) The O-H BDE of 94 kcal/mol (and the BDFE of 88 kcal/mol) is lower than that of the porphyrin complex, but it is still a tantalizing and perhaps a more accessible goal. Regenerating the *trans*-dioxovanadium(V) state via one-electron oxidation followed by a deprotonation would require a potential of 0.8 V and a pK_a of 16.3. Furthermore, the *cis* isomer of *trans*-dioxovanadium(V) with corrin is lower in energy by 30 kcal/mol, which suggests that generating the initial dioxovanadium(V) catalyst will require a route that kinetically favors the *trans* isomer.

A.



B.

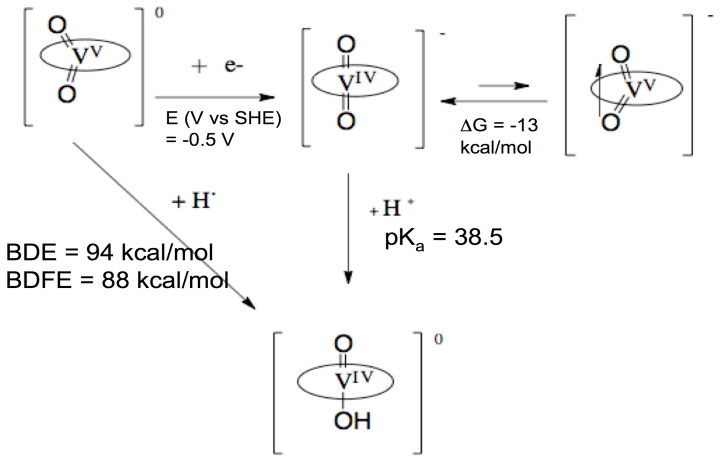


Figure 9.7 Thermochemistry of *trans*-dioxovanadium(V) corrin compounds. **(A)** Structure of the corrin macrocycle. **(B)** The reduction potential (E), free energy (*G*), pK_a, bond dissociation enthalpy (BDE), and bond dissociation free energy (BDFE) relating oxohydroxovanadium(IV) from *trans*-dioxovanadium(V) are shown above.

Another possible candidate to replace THP is the macrocycle is tetra-(N,N-dimethyl)imidazolium porphyrin (TDMImP) shown in **Figure 9.8**, which we have used previously in manganese complexes (20). We hoped that the electron-withdrawing groups of this ligand would decrease the hydroxo pK_a's during the protonation step and make regeneration of *trans*-dioxovanadium(V) more facile. We find that in the case of TDMImP, the BDE of *trans*-dioxovanadium(V) increases to 104 kcal/mol, which is comparable to that of methane, but too large for the compound to be stable. In **Table 9.1**, we compare the BDE and BDFE for the *trans*-dioxovanadium(V) compounds with three different ligands. All of the complexes have BDE's that are capable of activating strong CH bonds, which make these vanadium catalysts appealing candidates.

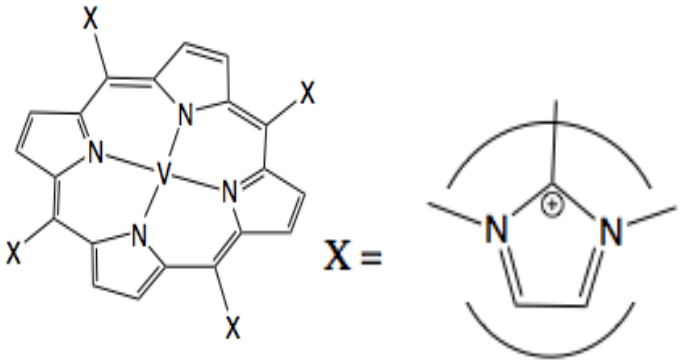


Figure 9.8 Structure of TDMImP macrocycle.

Table 9.1 BDE and BDFE of *trans*-dioxovanadium(V) with different macrocyclic ligands.

Ligand	BDE (kcal/mol)	BDFE (kcal/mol)
THP	103	97
Corrin	94	88
TDMImP	104	97

Fate of alkyl radical after hydrogen atom transfer (HAT)

Figure 9.9 shows the relative free energies (*G*) of the transition states, intermediates, and products resulting from *trans*-dioxovanadium(V) porphyrin complex abstracting a hydrogen from methane. Due to the high O-H BDE, the calculated activation free energy is a mere 20.3 kcal/mol. Following C-H activation, this reaction can proceed along two routes: rebound of the alkyl radical onto the hydroxo or the radical escape. In the present vanadium case, the barrier to methyl rebound is much higher than the energies of the radical escape products (18.8 kcal/mol vs. -2.6 kcal/mol). All the energies shown in **Figure 9.9** are for molecules in the singlet state (S=0) for consistency. We determined that the triplet state is lower in energy for the rebound transition state, but only by 0.6 kcal/mol. These energetics indicate that we would observe the formation of alkyl radicals and oxohydroxovanadium(IV) rather than methanol and oxovanadium(III). The alkyl radicals could further react with each other or different species to get a variety of products. The nucleophilic *trans*-dioxovanadium(V) compounds are therefore an interesting target based both on their predicted ability to activate alkanes and the possibility of accessing carbon-carbon bonds or other novel functionalization products of radical escape.

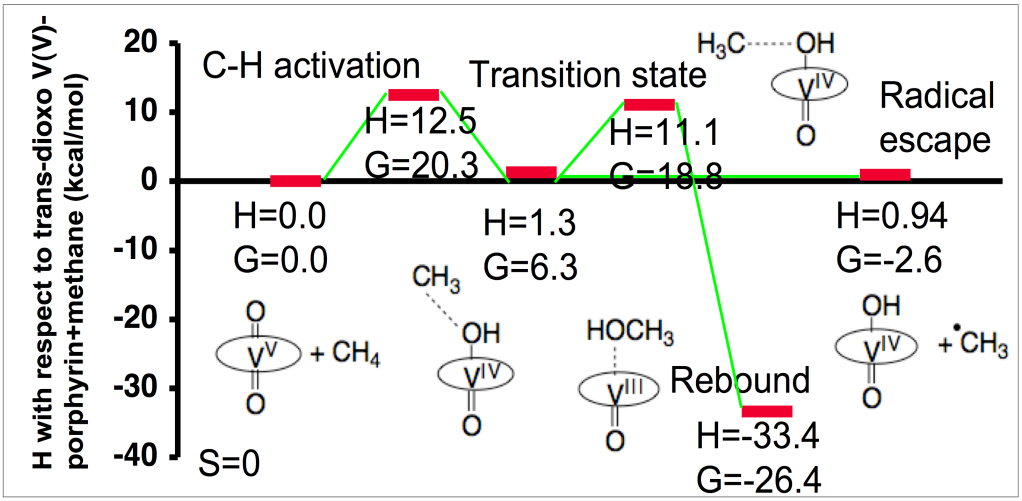


Figure 9.9 Enthalpy (*H*) and free energy (*G*) of the intermediates, transition states, and products resulting from methane activation by *trans*-dioxovanadium(V) porphyrin. *H* and *G* (kcal/mol) are shown with respect to *trans*-dioxovanadium(V) and methane as the reference (H=0.0, G=0.0).

CONCLUSIONS

Here, we have presented convincing evidence that the *trans*-dioxovanadium(V) compounds would be effective in activating hydrocarbon bonds. In complex with three different ligands, THP, corrin, and TDMImP, this compound's OH bond dissociation enthalpy is 94-104 kcal/mol and comparable to that of methane. We have also proposed a possible pathway to synthesize this compound from oxohydroxovanadium(IV). These vanadium complexes are more appealing than the previously characterized manganese ones because the vanadium complexes are more reactive, and they are more likely to produce alkyl radicals that can perform interesting chemistry rather than alcohols. Hence, we anticipate that these vanadium catalysts will play an important role in the future of the hydrocarbon bond activation field.

REFERENCES

1. Voluntary Reporting of Greenhouse Gases Program (Voluntary Reporting of Greenhouse Gases Program Fuel Carbon Dioxide Emission Coefficients), Washington, D.C.: U.S. Energy Information Administration; 2011.
2. Blanksby SJ, Ellison GB (2003) Bond dissociation energies of organic molecules. Acc Chem Res **36**:255-263.
3. Liu W, Huang XY, Cheng MJ, Nielsen RJ, Goddard WA, III, Groves JT (2012) Oxidative aliphatic C-H fluorination with fluoride ion catalyzed by a manganese porphyrin. Science **337**:1322-13255.
4. Liu W, Groves JT (2010) Manganese Porphyrins Catalyze Selective C-H Bond Halogenations. J Am Chem Soc **132**:12847-12849.
5. Golisz SR, Gunnoe TB, Goddard WA, III, Groves JT, Periana RA (2011) Chemistry in the center for catalytic hydrocarbon functionalization: An energy frontier research center. Catal Lett **141**:213-221.
6. Jaguar, version 7.9. (2012). Schrodinger, LLC, New York, NY.
7. Becke AD (1993) Density-functional thermochemistry 3. The role of exact exchange. J Chem Phys **98**:5648-5652.
8. Becke AD (1993) A new mixing of Hartree-Fock and local density-functional theories. J Chem Phys **98**:1372-1377.
9. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A: At Mol Opt Phys **38**:3098-3100.
10. Lee CT, Yang WT, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B: Condens Matter **37**:785-789.
11. Clark T, Chandrasekhar J, Spitznagel GW, Schleyer PV (1983) Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F. J Comput Chem **4**:294-301.
12. Frisch MJ, Pople JA, Binkley JS (1984) Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. J Chem Phys **80**:3265-3269.

13. Krishnan R, Binkley JS, Seeger R, Pople JA (1980) Self-consistent molecular orbital methods 20. Basis set for correlated wave-functions. J Chem Phys **72**:650-654.
14. McLean AD, Chandler GS (1980) Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11-18. J Chem Phys **72**:5639-5648.
15. Hay PJ, Wadt WR (1985) Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. J Chem Phys **82**:270-283.
16. Zhao Y, Truhlar DG (2008) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theor Chem Acc **120**:215-241.
17. Tannor DJ, Marten B, Murphy R, Friesner RA, Sitkoff D, Nicholls A, Ringnalda M, Goddard WA, III, Honig B (1994) Accurate first principles calculation of molecular charge distributions and solvation energies from ab initio quantum mechanics and continuum dielectric theory. J Am Chem Soc **116**:11875-11882.
18. Marten B, Kim K, Cortis C, Friesner RA, Murphy RB, Ringnalda MN, Sitkoff D, Honig B (1996) New model for calculation of solvation free energies: Correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. J Phys Chem **100**:11775-11788.
19. Kelly CP, Cramer CJ, Truhlar DG (2006) Aqueous solvation free energies of ions and ion-water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. J Phys Chem B **110**:16066-16081.
20. Umile TP, Wang D, Groves JT (2011) Dissection of the mechanism of manganese porphyrin-catalyzed chlorine dioxide generation. Inorg Chem **50**:10353-10362.